EXTREMAL BEHAVIOR OF THE SURFACE OZONE OVER THE CONTINENTAL U.S.

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Pakawat Phalitnonkiat

January 2017

EXTREMAL BEHAVIOR OF THE SURFACE OZONE OVER THE CONTINENTAL U.S.

Pakawat Phalitnonkiat, Ph.D.

Cornell University 2017

The ground level ozone concentration over the continental United States is analyzed from the point of view of modern Extreme Value Theory using ozone data from the Clean Air Status and Trends Network (CASTNET) at 25 measurement sites. First, we estimate the changes in ozone means according to the NO_X SIP call policy implemented during 2000 in Northeastern U.S. The most significant change time in ozone mean is estimated in both parametric and non-parametric ways. The change in variability is also estimated but the results are not as significant as the change in mean. The results show that the policy is effective in reducing the ozone means within 2-4 years. Thus, to analyze the effects of the policy in the extreme sense, the ozone data is divided into two climate schemes. Then, the Generalized Pareto Distribution is fit to extremes of the ozone concentration by using a combination of maximum likelihood estimates (MLEs) and Hill estimates. The data is transformed prior to extreme value analysis and data in the right tail is separated from that in the middle part of the distribution. This analysis is compared to current approaches by using synthetic data. Under a variety of conditions the procedure using the MLE approach is likely to underestimate the tail of the distribution. The analysis shows that at some CASTNET locations the ozone probability distribution is not exponentially bounded, and thus can be characterized as heavy tailed. The ozone tail distributions become heavier following the NO_X SIP call at most of the sites with heavy tails prior to this call.

In the final part, we study the extreme dependence between temperature and ozone. We also use simulated data from existing models in the study to verify the model correctness. These models are widely used in climate analysis; however, in terms of extremes, the models usually do not well represent the relationship between temperature and ozone compared to the CASTNET measurements. The models tend to have higher extreme dependence than those from CASTNET data, and hence, we should be careful when using data from these models in extreme studies.

BIOGRAPHICAL SKETCH

Pakawat Phalitnonkiat was born and raised in Chonburi, Thailand. He was a finalist for the Representatives of Thailand for 45^{th} and 46^{th} International Mathematical Olympiad. Then he received a scholarship from Institute for the Promotion of Teaching Science and Technology (IPST) to study Bachelor degree aboard.

He came to the U.S. in 2006 and studied at Western Reserve Academy for one year as a preparation for college. By that time, he was an USAMO Qualifier. After that he obtained his bachelor's degree in mathematics and economics minor in Computer Science with Magna Cum Laude with distinct in all subjects from Cornell University. After the graduation, he continues his academic life at Cornell University pursuing his Ph.D. program in Applied Mathematics under the supervision of professor Gennady Samorodnitsky. His research was conducted on extreme value theory applied to climate. After completing his Ph.D. program, he will be working in the area of data analysis.

To my dad, mom, grandma, grandpa and sister

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my remarkable advisor, professor Gennady Samorodnitsky for his patience, friendship, guidance and supports. Thank him for accepting me as a part of his research group and being a great mentor. Without his guidance and persistent help, this dissertation would not have been possible. I would also like to thank professor Peter G.M. Hess, professor Mircea D. Grigoriu, and professor Tara S. Holm, who served on my dissertation committees, for their help over all the years at Cornell University. I also thank Wenxiu Sun for her collaborations on all of the projects.

In this very special moment, I give my appreciation to my family for their unceasing encouragement and extraordinary support. I give a special thank to Kunlaya Soiaporn for her help in many statistical techniques, as well as other academic and non-academic supports. I also thank Nongnuch Athiphunamphai and Chinawat Isradisaikul for all the supports, advices, and patience.

I would like to thank my fellow students at Center for Applied Mathematics for all good conversations and coffee hours which have made our office a warm place to work. Moreover, super thanks to all my Thai friends at Cornell. I could not name you all but I do remember everyone and everything we did together. You are always there for me through my good and bad times. Last but not least, I would like to acknowledge the Environmental Protection Agency (EPA) for the financial supports and giving me the great opportunity for being a part of Cornell University.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Any major amount of exposure to surface ozone, either chronic or acute, has been associated with increased risk of death from cardiovascular and respiratory causes (Bell et al., (2004)[4]; Jerrett et al., (2009)[26]) and damage to agriculture estimated in 2000 at $14-$26 billion per year (van Dingenen et al., (2009)[50]). Ozone, a secondary pollutant, is generated through the oxidation of volatile organic carbons, carbon monoxide and methane in the presence of NO_X and sunlight. The analysis from Duncan et al., (2010)[10] shows that across most of the continental U.S. ozone production is NO_X limited. Ozone shows significant temporal variability within the continental boundary layer across a wide range of timescales with pronounced inter-annual, seasonal, daily and hourly variability. Since 1971, ozone has been regulated in the U.S. under the Clean Air Act. In 1997 the NAAQS (National Ambient Air Quality Standard) for ozone was set to a maximum daily 8-hour average ozone (MDA8 O3) of 84 ppb. To meet this standard, the NO_X State Implementation Plan (SIP) was implemented for controlling NO_X emissions in 22 states in the Eastern U.S.

Summertime ozone concentrations over the U.S. can be thought to be comprised of two portions: a portion amendable by U.S. emissions controls and a background value not sensitive to these controls (e.g., Fiore at al., (2014)[15]). Summertime background values range from 25 to 40 ppb over high-altitude western sites to 20-30 ppb over the Eastern U.S. (Fiore at al., (2014)[15]). Extreme ozone pollution concentrations above background levels can sometimes be asso-

ciated with tropopause folds (Lin et al., (2012)[29]) or with wildfires (e.g., Jaffe et al., (2012)[25], but more frequently with an intensification of pollution events in association with particular meteorological conditions. Intense pollution events are often associated with warm stagnant conditions over the East Coast of the U.S. (Hegarty et al., (2007)[20]; Vukovich, (1995)[53]) and often occur with weak, slow-moving and persistent high-pressure systems (Logan, (1989)[30]). The 2003 European heat wave and the associated strong stagnation is an example of a severe meteorological event accompanied by dangerous ozone concentrations (Vautard et al., (2005)[52]; Vautard et al., (2007)[51]; Guerova et al., (2007)[18]; Solberg et al., (2008)[46]). Summertime ozone concentrations are highly correlated with temperature (Brown-steiner et al., (2015)[6] and references therein). This is at least partly explained by the correlation between temperature, air stagnation, and solar radiation (e.g., Jacob et al., (2009)[24]. There is a number of feedbacks between the severity and duration of heat waves and surface pollution concentrations including changes in natural emissions particularly those from fire and vegetation (Fang et al., (1996)[14]; Guenther et al., (2006)[17]; Guenther et al., (2012)[16]) and changes in stomata ozone uptake with its impact on dry deposition (Vautard et al., (2005)[52]; Solberg et al., (2008)[46]). On the other hand, the tropospheric chemical system appears to be highly buffered (Shindell et al., (2003)[45]; Stevenson et al., (2006)[48]) so that the response of ozone to changes in chemical or meteorological forcing may be weaker than expected.

Temperatures (IPCC, (2013)[23]), the frequency and severity of heat waves (Russo et al., (2014)[43]), and in many parts of the globe air stagnation days (Horton et al., (2014)[22]) are expected to increase during this next century. However, NO_X emissions, in many locations the limiting ozone precursor, are expected to decrease. A number of studies have suggested that climate change alone will pref-

erentially increase the frequency of extreme pollution events on the high end of the cumulative probability distribution (Mickley et al., (2004)[31]; Hogrefe et al., (2004)[21]; Tagaris et al., (2007)[49]; Wu et al., (2008)[55]) although other studies do not show this (Murazaki et al., (2006)[33]; Lin et al., (2008)[28]; Rieder et al., (2015)[40]). The projected future decrease in NO_X emissions, on the other hand, is expected to decrease both the mean and the 90^{th} percentile ozone concentrations (Rieder et al., (2015)[40]) over the Eastern U.S. with the 90^{th} percentile more sensitive to NO_X changes than the mean concentration.

Many researchers suggested that the ozone has been decreased due to the NO_X SIP implementation (e.g., Rieder et al., (2013)[41]). In Chapter 2, we verify the impacts of the NO_X SIP implementation on the ozone distributions by using change-point detection methodology. The NO_X SIP implementation has been regulated only in the eastern U.S. but to measure the effectiveness of the implementation, we compare the results to other areas over the continental U.S. as well. In our study, we use CASTNET (The Clean Air Status and Trends Network) ozone and temperature data over the continental U.S. The goal of study is to find the time frames when ozone means have shifted down and determine if the dispersions of the ozone distribution have changed. This study also looks at the same properties of temperature distributions to determine if the changes in ozone and temperature are concurrent. The results from this chapter provide fundamental information and a rough picture of the temperature and ozone distributions which will be useful for the next two chapters.

In Chapter 3, we further characterize ozone extremes over the continental U.S. using Extreme Value Theory (EVT) which provides a useful mathematical framework to examine the tail of temperature and ozone distributions. We usually refer

the tail distributions to as the right end of the distributions (Coles, (2001)[8]). The results are used to characterize the geography of the extreme behavior of measured ozone concentrations. We achieve this goal by refining the approach in Rieder et al., (2013)[41] via specialized methods for estimating tail properties of the ozone distribution. These methods are tested on synthetic data and applied subsequently to analyze CASTNET ozone data over the continental U.S.

There are studies suggesting a suppression in ozone occurs when temperature becomes higher (e.g., Shen et al., (2016)[44]). This interests us in the relationship between extreme temperature and extreme ozone. Therefore, in Chapter 4, we study the dependency in extremes. To achieve this, we introduce certain notions in multivariate extreme value theory and develop tools for their applications. We also increase the data pool by importing more data from three different climate models that are widely used in climate studies. The first model is specified dynamics model (SDM) which uses data assimilation from meteorological fields. The model is claimed to be robust and good for many climate studies. The other two models are general circulation models (GCMs). One is for current climate scheme (2000-2025) and the other one is projected for future climate scheme (2100-2125). In the end, we found out that the models disagree with the observed data from CASTNET in the extreme relationship between temperature and ozone. This analysis suggests that when we use the simulations from these models to study extreme climates, we have to be cautious about the details of the models.

In summary, this dissertation presents the effects of NO_X SIP implementation by showing the changepoint in ozone and temperature locations and scales, and the increases in tail indices from most CASTNET sites. Moreover, we measure the relationship between extreme temperature and extreme ozone and it turns out to

be that the extreme dependency is much different than what we expect from the models commonly used by climatologists. This discrepancy is significant enough so that it is worth mentioning for not falling into these caveats.

CHAPTER 2

CHANGEPOINT DETECTION: A NEW ERA OF SURFACE OZONE CONCENTRATIONS

2.1 Introduction

Changepoints are instances at which statistics of time series exhibit notable variations. Changepoints can also be viewed as changes in parameters of the distributions. Sometimes the changes are hard to verify, yet the changepoint-free time series is even harder to confirm. One of the applications in changepoint detection is to evaluate the impacts from the changes of policies or some major events (Alexandersson (1986)[2]; Easterling and Peterson (1995)[11]; Apollonio et al., (2016)[3]). There are a lot of studies that are devoted to deep analysis of both theory and applications of the methodology. For example, Hawkins and Deng (2010)[19] uses non-parametric methods to find changepoints in the mean of time series. The study suggested two approaches of finding changepoints. The first approach assumes that the length of the time series is fixed and we have to find the changepoints at once. The second approach can take real time inputs and compute the changepoints on the fly. Both approaches are quite similar in terms of that they calculate the test statistics and flag a point if the test statistic is higher than some threshold.

For this study, we will only use the fixed-length approach because we are not analyzing the real time data. Our main goal is to analyze the results of NO_X SIP call policy originally imposed at the beginning of 2000's in New England with the goal of regulating the nitrogen oxide emissions which would also reduce ozone level in the process. The policy has expanded the effective areas to the southern New England later in 2003. There are a lot of studies focusing on the decreases in means

of ozone after the policy has been introduced (e.g., Bloomer et al., (2009)[5]; Butler et al., (2011)[7]; Aleksic et al., (2013)[1]). Although many studies have reported the changes in ozone means, they did not specifically locate the changes in time (Aleksic et al, (2013)[1]). This is our opportunity to verify those studies and the results from this chapter will be basis for the next chapters as well. In this chapter, we will introduce some existing methods to detect changepoints. Since different methods have their own strong and weak points, we would like to compare the efficiency of these methods on climate data as well.

The composition of this chapter is as follows: Section 2.2 explains how the data is obtained and what needs to be done before using it in analysis; Section 2.3 explains the methods that we use to find changepoints and also estimate test statistics to decide if the results are significant enough; Section 2.4 shows the results when we apply the changepoint detection to observed data; Section 2.5 discusses the results and how we can use them in future studies.

2.2 Data

Ozone measurements from the CASTNET (www.epa.gov/ castnet) observational network are used to search for ozone changepoints at 25 sites throughout the continental U.S. The hourly CASTNET data is converted into daily data by calculating the maximum daily 8-hour average of ozone concentration (MDA8). We use the summertime (June, July and August–JJA) ozone measurements between 1992 and 2013. Table 2.1 shows the basic information from each CASTNET site we used in our study.

Table 2.1: Means and standard deviations of temperature ($^{\circ}C$) and ozone (ppb) of each CASTNET site that we use in our analysis. The data is collected from 1992 to 2013.

Site name	Longitude	Latitude	State	Temperature ($°C$)		Ozone (ppb)	
				Mean	SD	Mean	SD
Ashland	68.4W	46.6N	ME	21.296	4.173	35.307	10.149
Howland	68.7W	45.2N	ME	22.703	4.169	37.502	12.078
Woodstock	71.7W	43.9N	NH	21.968	6.175	36.500	10.932
Connecticut Hill	76.7W	42.4N	NY	21.932	3.910	50.860	13.115
Penn State	77.9W	40.7N	PA	24.289	3.608	54.551	15.017
Parsons	79.7W	39.1N	WV	24.317	3.140	51.590	12.544
Edgar Evins	85.7W	36N	TN	27.090	2.955	50.189	12.762
Georgia Station	84.4W	33.2N	GA	28.477	2.822	52.929	17.840
Sand Mountain	86W	34.3N	AL	27.511	2.823	54.508	14.895
Candor	79.8W	35.3N	NC	28.172	3.305	53.201	16.126
Coweeta	83.4W	35.1N	NC	25.239	2.772	41.547	12.488
Cranberry	82W	36.1N	NC	20.498	2.602	54.008	11.374
Beaufort	76.6W	34.9N	NC	27.733	2.315	44.735	14.153
Prince Edward	78.3W	37.2N	VA	27.700	3.590	51.716	13.689
Shenandoah NP	78.4W	38.5N	VA	19.954	3.152	58.695	12.110
Horton Station	80.6W	37.3N	VA	22.345	2.958	56.768	12.229
Sumatra	85W	30.1N	FL	29.576	2.336	36.986	13.386

Table 2.1: (Continued).

Site name	Longitude	Latitude	State	Temperature ($°C$) Mean	Temperature ($°C$) SD	Ozone (ppb) Mean	Ozone (ppb) SD
Bondville	88.4W	40.1N	IL	26.395	3.616	53.386	13.844
Perkinstown	90.6W	45.2N	WI	21.893	3.936	42.779	11.564
Oxford	84.8W	39.5N	OH	26.101	3.541	56.583	14.512
Glacier NP	114W	48.5N	MT	21.577	5.524	38.205	9.090
Chiricahua NM	109.4W	32N	AZ	28.401	3.374	54.044	8.328
Grand Canyon NP	112.2W	36.1N	AZ	24.397	3.519	56.534	7.651
Gothic	107W	40N	CO	16.803	3.671	53.426	7.333
Lassen Volcanic NP	121.6W	40.5N	CA	20.687	4.732	54.195	10.255

To search for changepoints, it is necessary to transform the daily observations in order to reduce the seasonal variability within a JJA period (i.e., to prevent significant distributional changes from day to day). We call this method deseasonalization. First, we denote each data point by $x_{y,d}$ for the day d in year y. We use the summertime (June, July and August–JJA) ozone measurements between 1992 and 2013. Thus, there are only 92 days for each year and we refer $d = 1$ to June 1^{st}, and refer $d = 92$ to August 31^{st} of a particular year.

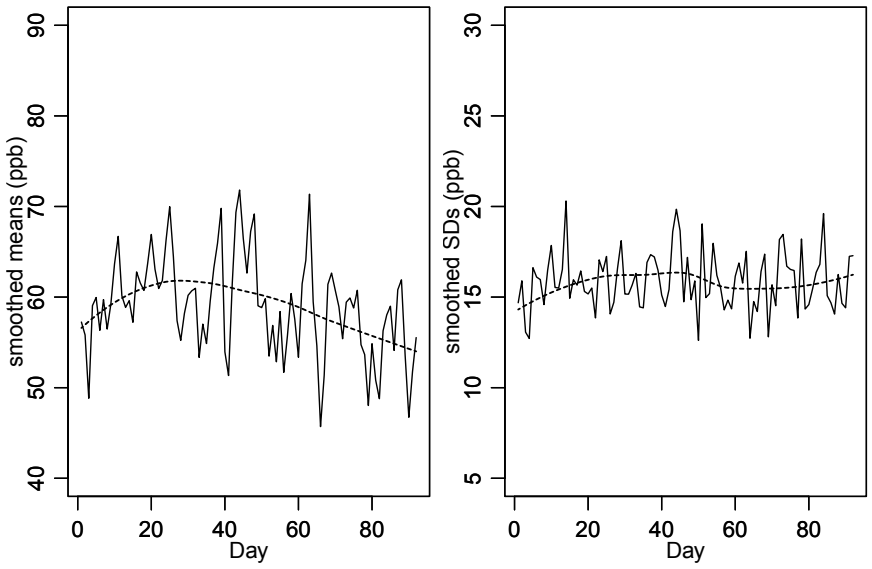


Figure 2.1: Ozone averaged over 22 years (1992-2013) (left) and ozone standard deviations (right) for each day in one year from a selected station (Penn state (PA)). Day 1 refers to June 1 of each year and Day 92 refers to August 31 of each year. Dotted curves are smoothed ozone means (left) and smoothed ozone standard deviations (right).

Deseasonalization (Minimizing the intra-annual variability): To minimize the seasonal effects that occur within a year, we calculate, for each day d, the mean and the standard deviation of the measured ozone concentrations for that day, over

all available years. That is, we compute, for each day d,

$$m_d = \frac{1}{Y}\sum_{y=1}^{Y} x_{y,d}, \quad sd_d = \sqrt{\frac{1}{Y}\sum_{y=1}^{Y}(x_{y,d} - m_d)^2}\,, \tag{2.1}$$

where $Y = 22$ is the number of years in the data. In principle, the intention is, then, to normalize the observations according to the rule

$$\hat{x}_{y,d}^{DS} = \frac{x_{y,d} - m_d}{sd_d}\,. \tag{2.2}$$

We sometimes refer this to normalized scale data later in the texts. However, it is clear that the sample size $Y = 22$ is really very modest, and the resulting variability from day to day in the estimates in (2.1) will be, in part, due to the noise. Therefore, prior to performing the normalization suggested in (2.2) we, first, put the estimated daily means and standard deviations through a smoothing procedure using a local polynomial regression. In order not to overburden the notation, we will still use the notation m_d and sd_d for the smoothed values of the estimates. As an example, the raw estimates and the smoothed estimates of a particular station (in our example, we use Penn State (PA)) are presented in Figure 2.1. We notice from the left panel of Figure 2.1 that there is a rather pronounced seasonal cycle in ozone concentrations during JJA rendering the original data non-stationary.

2.3 Methodology

In this section, we are presenting three methods of finding changepoints. Two of them can be used to check if means of two groups are the same. For both methods, we also assume that the data are iid. For this chapter, we only find the most significant changepoint of the time series. We could find multiple changepoints by iteratively using the methodology on the sub-series. The first method

assumes the data are drawn from normal distributions which is somewhat hard to justify in practice. The other method is a non-parametric way to determine the changepoint of a time series without an assumption of normality. However, the independence is still assumed. The last method is for finding a changepoint in variability. This method is also non-parametric and uses a similar technique as when we find changepoints in the mean.

2.3.1 Parametric Case

In this case, we assume that the data are independent and drawn from normal distributions. We can use a test that looks like two-sample t-test to test if two samples have the same means. That is, we setup our hypotheses as

$$H_0 : X_i \sim N(\mu_0, \sigma^2) \text{ for } 1 \leq i \leq n; \tag{2.3}$$

and for some c,

$$H_1 : \begin{cases} X_i \sim N(\mu_1, \sigma^2) & \text{for } 1 \leq i \leq c \\ X_i \sim N(\mu_2, \sigma^2) & \text{for } c+1 \leq i \leq n \end{cases}, \tag{2.4}$$

where $\mu_0, \mu_1, \mu_2, \sigma, c$ are unknown. Since we do not know an exact time that a changepoint may occur, we calculate the test statistic for each time c by using the following formula:

$$\hat{P}_c = \frac{\bar{X}_1 - \bar{X}_2}{s_{12}\sqrt{c^{-1} + (n-c)^{-1}}}, \tag{2.5}$$

where $\bar{X}_1 = \frac{1}{c}\sum_{i=1}^{c} X_i$ and $\bar{X}_2 = \frac{1}{n-c}\sum_{i=c+1}^{n} X_i$ are sample means before and after the time c, respectively, and

$$s_{12} = \left(\frac{(c-1)s_1^2 + (n-c-1)s_2^2}{n-2} \right)^{1/2} \tag{2.6}$$

is the pooled standard deviation of two samples with s_1, s_2 being sample standard deviations before and after time c. We are finding the value c that maximizes $|\hat{P}_c|$ and this refers to the most probable changepoint at location c, however, we try to avoid changepoint near the boundaries, so we could cut about 10% of each boundary out. That is, we calculate

$$P = \max_{\frac{n}{10} \leq c \leq \frac{9n}{10}} |\hat{P}_c|. \tag{2.7}$$

The test statistic P can show if the time series has a changepoint by using critical values which will be discussed in Section 2.3.3. If a changepoint occurs, the argmax in (2.5) is the location of the most significant change time of the time series.

2.3.2 Non-parametric Case

The assumption of the data being Gaussian in the parametric case is a strong assumption and may not be realistic. Instead, we can use a non-parametric variant to help us avoid this issue. We use Mann-Whitney statistic to detect a changepoint by maximizing it over time c and if a change occurs, then the argmax c is most significant location time. We define rank of X_i by $r_i = \sum_{t=1}^{n} \mathbf{1}_{\{X_t \geq X_i\}}$. Using ranks instead of observed values usually makes a statistical test mores robust in the sense that when the parametric assumption is violated, the test can still carry on with almost the same efficiency if the sample size is large enough, Reeves et al., (2007)[37]. Now, we consider the Mann-Whitney statistic $W_c = \sum_{i=1}^{c} r_i$, then calculate the mean and variance of W_c:

$$\mathbb{E}(W_c) = \frac{c(n+1)}{2} \text{ and } Var(W_c) = \frac{c(n-c)(n+1)}{12}. \tag{2.8}$$

From the asymptotic normality of Mann-Whitney statistics, if we standardize the test statistic W_c, it will be asymptotically Gaussian with mean 0 and variance

1 as c and $n-c$ go to infinity (Hawkins and Deng, (2010)[19]). That is, we have the asymptotic normality of the test statistic

$$W'_c := \frac{\sum_{i=1}^{c} r_i - c(n+1)/2}{\sqrt{c(n-c)(n+1)/12}} \sim N(0,1) \text{ as } c \to \infty \text{ and } n-c \to \infty, \tag{2.9}$$

where r_i is the rank of X_i. The presence of a changepoint and the time of its occurrence can be detected by maximizing the statistics $|W'_c|$ over c. That is, we use

$$W = \max_{\frac{n}{10} \le c \le \frac{9n}{10}} \left| \sqrt{12} \cdot \frac{\sum_{i=1}^{c} r_i - c(n+1)/2}{\sqrt{c(n-c)(n+1)}} \right| \tag{2.10}$$

as our test statistic to determine a change in mean of a time series. It is worth to note that Hawkins (2010)[19] found out that when the sample size is large enough, the test statistics P and W effectively have the same statistical properties. This gives us almost identical results for our study, see Table 2.3.

Based on the asymptotic normality of the Mann-Whitney test statistics (Ross, (2011)[42]), we can use the same technique to find a changepoint in dispersion. The method is still non-parametric but instead of using the rank-sum W_c, we measure the deviation of rank from its expected value. The test is called Mood test, first developed by Mood, (1954)[32]. The main idea of this test comes from the observation that if the null hypothesis holds, i.e., no change in dispersion in the time series, then with iid property, the expected value of rank r_i is $(n+1)/2$. Mood test tries to measure the deviation in rank of each point from its expected value by using the sum of the rank deviations squared:

$$M_c := \sum_{i=1}^{c} \left(r_i - \frac{n+1}{2} \right)^2. \tag{2.11}$$

Next, we calculate its expected value and variance of M_c:

$$\mathbb{E}(M_c) = \frac{c(n^2-1)}{12} \text{ and } Var(M_c) = c(n-c)(n-2)(n+1)(n+2)/180. \tag{2.12}$$

Using asymptotic normality in Mann-Whitney statistic, we can define

$$M'_c = \frac{\sum_{i=1}^{c}(r_i - (n+1)/2)^2 - c(n^2-1)/12}{\sqrt{c(n-c)(n-2)(n+1)(n+2)/180}} \tag{2.13}$$

as our test statistic. Therefore, we have $M'_c \sim N(0,1)$ as $c, (n-c) \to \infty$. We can determine the presence of the change and its occurrence, if exists, by maximizing the test statistics $|M'_c|$ over c:

$$M = \max_{\frac{n}{10} \le c \le \frac{9n}{10}} \left| \sqrt{180} \cdot \frac{\sum_{i=1}^{c}(r_i - (n+1)/2)^2 - c(n^2-1)/12}{\sqrt{c(n-c)(n-2)(n+1)(n+2)}} \right|, \tag{2.14}$$

and if a changepoint occurs, then the argmax is its occurrence time.

2.3.3 Estimating the critical levels

There are published critical values for the test statistics we introduced above, however, the results are inconsistent (e.g., Reeves et al., (2007)[37], Hawkins and Deng, (2010)[19]). Thus, we will perform some simulations to estimate the critical values again. For parametric case (Section 2.3.1), if we know the change time c, then the test is essentially two-sample t-test. However, in this case, the time c is unknown, so we have to simulate the data to find critical values for our test statistics. For non-parametric case, we also simulate the data to find critical values for test statistics W and M. The results of the critical values are summarized in the Table 2.2.

We simulate data sets with different numbers of points n = 50, 75, 100, 200, 500, 1000, 2000 and 4000. Then we repeat the simulation 1000 times and estimate the critical values of the test statistics. Table 2.2 summarizes critical values at different confident levels and with different number of data points n.

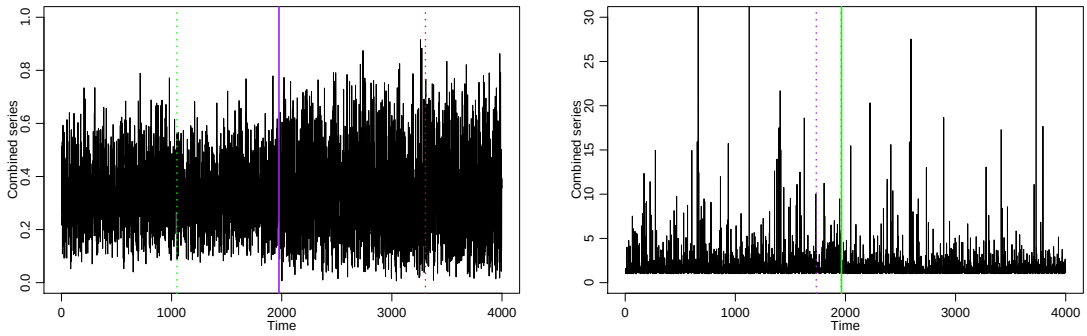
Table 2.2: Critical values at different confident levels for parametric and non-parametric methods. The values are also dependent on the length of the time series (n).

Test statistics	Parametric		Non-parametric			
	Mean P		Mean W		Variability M	
n	95%	99%	95%	99%	95%	99%
50	3.447	4.105	3.026	3.470	3.079	3.509
75	3.421	3.983	3.113	3.559	3.125	3.616
100	3.414	3.940	3.179	3.612	3.155	3.631
200	3.445	3.941	3.280	3.692	3.254	3.671
500	3.479	3.947	3.377	3.857	3.260	3.733
1000	3.513	3.985	3.437	3.919	3.273	3.748
2000	3.564	4.056	3.486	3.927	3.314	3.810
4000	3.644	4.058	3.540	3.980	3.338	3.897

Figure 2.2 shows examples of using different methods to detect changepoints in means and variability. We construct two examples by generating two time series of length $N = 2000$ from some distributions, call them $X^{(1)}$ and $X^{(2)}$, then append them together. For the first example, let $Beta(\sigma_1, \sigma_2)$ be beta distribution with shape parameters σ_1, σ_2. In Figure 2.2a, we generate $X^{(1)} \sim Beta(4, 8)$ and $X^{(2)} \sim Beta(2, 4)$. Notice that both sub series have the same means of $\frac{1}{3}$ but different variance. The first series $X^{(1)}$ has variance $\frac{2}{117}$ and the second series $X^{(2)}$ has variance $\frac{2}{63}$. When we are finding changepoints in means, both parametric and non-parametric methods do not detect the changepoint in mean due to lacking of significant test statistics. However, the non-parametric method for finding changepoints in variability can detect this. The location of changepoint is $c = 1975$ with test statistic $M = 15.544$ (represented by the solid purple line).

We also provide another example when the means and variance are equal even though they are from different distributions. Figure 2.2b shows the case of $X^{(1)} \sim GPD(1/2, 1, 1/2)$ and $X^{(2)} \sim GPD(1/2, 1, 7/20)$, where $GPD(\xi, \mu, \sigma)$ denotes the generalized Pareto distribution with shape parameter ξ, location pa-

rameter μ and scale parameter σ. From Figure 2.2b, the non-parametric method finds a changepoint in mean at $c = 1964$ with test statistic $W = 8.227$, and this is represented by the green line.



(a) $Beta(4, 8)$ and $Beta(2, 4)$ (b) GPD(1/2,1,1/2) and GPD(1/2,1,7/20)

Figure 2.2: (a) Two time series combined together $(X^{(1)}, X^{(2)})$, where $X^{(1)} \sim Beta(4, 8)$ and $X^{(2)} \sim Beta(2, 4)$. The means from each sub series are equal but with different variances. The optimal c from parametric method is $c = 3305$ with test statistic $P = 1.89$ (shown by dotted red line), and the optimal c from non-parametric method (finding changes in means) is $c = 1048$ with test statistic $W = 2.09$ (shown by dotted green line). The non-parametric method for finding changepoints in variance detects a chagnepoint at $c = 1975$ with test statistics $M = 15.54$ (shown by solid purple line). (b) Two time series combined together using General Pareto distribution (GPD) $(X^{(1)}, X^{(2)})$, where $X^{(1)} \sim GPD(1/2, 1, 1/2)$ and $X^{(2)} \sim GPD(1/2, 1, 7/20)$. Both series do not have the same means or variance. The parametric test fails to give enough evidence to conclude that the change exists, however, the non-parametric test concludes that there is a change at $c = 1964$ with enough evidence of $W = 8.227$. The non-parametric method for variance gives optimal $c = 1719$ with test statistic $M = 3.15$ which does not provide enough evidence for a changepoint in variance (dotted purple line).

2.4 Detecting Changepoints in CASTNET Data

Since we would like to evaluate the results of NO_X SIP call in reducing the ozone levels, we expect the ozone level to start decreasing at some point after the policy

was in effect, especially CASTNET sites located in eastern U.S. To verify this, we first take out seasonal effects from the data during the JJA period. The method for deseasonalization is described in Section 2.2. Then we apply the changepoint detection methods to the ozone levels observed at the CASTNET sites. We apply the three methods and compare the results among them. See Table 2.3 for the changepoint and test statistics from each site and the last column shows the results from the non-parametric method for finding changepoints in variability. Note that we only consider summertime data, so there are 92 days for each year and the optimal indices c refer to the day in time series. For example, $c = 1$ refers to June 1, 1992, and $c = 93$ refers to June 1, 1993, etc. We should also emphasize that only the most significant changepoint is reported in this table. There could be multiple changepoints.

Since we use the data over 22 years, there are 2024 data points for each site (including some N/A data). We can see that most indices c from changepoints in mean of the sites located in eastern U.S. are reported to be around c = 990-1100. These indices refer to the years 2002-2004. However, this is not the case for the variability of ozone distributions. Only few sites have changed variability during 2002-2004 which is the time frame for change in ozone means. At 95% significance level, there are also some CASTNET sites (in New England and Western U.S.) that do not have enough evidence to conclude that the variability has changed and they are marked by *. Recall that the NO_X SIP call started being active during 2000 in New England and expanding the effective area to southern New England in 2003. Since ozone level is also affected by the reduction in nitrogen oxide emissions, these results show that there is some delay between the policy start date and the ozone reduction. Table 2.4 also shows the differences between means of ozone before and after the changepoint and the difference between them.

Table 2.3: Results from each method applied on deseasonalized ozone data from CASTNET sites. * is marked if the test statistic is lower than 95% critical values according to Table 2.2.

		Parametric		Non-parametric			
		Mean		Mean		Variability	
Site	State	Optimal c	P	Optimal c	W	Optimal c	M
Ashland	ME	1120	12.95	1121	12.48	826	2.01*
Howland	ME	1073	10.93	1128	10.31	353	2.27*
Woodstock	NH	1122	11.65	1123	11.48	353	2.33*
Connecticut Hill	NY	1050	14.97	1050	13.74	996	4.24
Penn State	PA	1003	13.64	1003	12.46	1038	5.04
Parsons	WV	1005	18.98	1005	17.27	757	6.22
Edgar Evins	TN	994	13.9	994	12.85	850	4.6
Georgia Station	GA	994	9.81	1881	9.26	1049	6.27
Sand Mountain	AL	994	15.08	994	14.03	957	5.95
Candor	NC	994	19.57	994	17.68	1244	7.13
Coweeta	NC	1018	18.58	1023	17.07	500	5.75
Cranberry	NC	1006	18.49	1006	17.3	1052	4.49
Beaufort	NC	1833	9.66	1833	9.18	777	7.06
Prince Edward	VA	1007	17.93	1007	16.64	1271	6.84
Shenandoah	VA	1102	18.24	1008	16.75	755	5.07
Horton Station	VA	1005	21.2	1005	19.19	315	3.75
Sumatra	FL	829	13.04	829	11.58	1346	4.98
Bondville	IL	1401	15.77	1401	14.51	1592	3.51
Perkinstown	WI	1456	6.05	1457	5.56	1355	4.09
Oxford	OH	1047	12.84	1047	11.95	754	3.05*
Glacier NP	MT	898	10.52	898	10.35	551	4.3
Chiricahua	AZ	204	5.61	204	5.64	819	2.19*
Grand Canyon	AZ	1486	6.83	1486	6.84	220	5.54
Gothic	CO	1536	10	1536	9.7	514	3.21*
Lassen Volcanic	CA	1540	6.19	1540	5.75	1528	4.01

Apart from western U.S., almost all of the sites show the decrease in ozone means. These are desirable outcomes that we would expect from the policy.

We also apply changepoint detection in mean to daily maximum temperature. For the test statistics P and W from equations (2.7), (2.10), if we consider all time indices c from 1 to n then the changes in temperature means occur around the boundary of the time series. When we take out the boundaries, the changepoints do not happen in the middle as the ozone data do. The results are reported in

Table 2.5. For southeastern and western U.S., temperature means have changed around mid 2005-2006. We can also see that only few CASTNET sites have some changes in variability of temperature distributions. It is worth mentioning that most sites have the magnitudes of the test statistics of temperature lower than those of ozone, i.e., $P_{O_3} > P_{Temp}$, and $W_{O_3} > W_{Temp}$. In fact, the temperature means are actually increasing. Table 2.6 shows the means and the difference of temperature of before and after the change time c based on the non-parametric method.

Table 2.4: Ozone means (ppb) for each CASTNET sites before and after the changepoint. Differences between before and after the changepoints are also shown. The change time c is based on the non-parametric method.

Site	State	Ozone means (ppb)		
		Before c ($\bar{X}_1$)	After c ($\bar{X}_2$)	Difference ($\bar{X}_2 - \bar{X}_1$)
Ashland	ME	37.81	32.12	-5.69
Howland	ME	40.15	34.17	-5.98
Woodstock	NH	38.93	33.17	-5.76
Connecticut Hill	NY	54.65	46.36	-8.29
Penn State	PA	59.17	50.27	-8.9
Parsons	WV	56.45	46.66	-9.7
Edgar Evins	TN	54.08	46.44	-7.64
Georgia Station	GA	53.92	39.15	-14.77
Sand Mountain	AL	59.54	49.92	-9.62
Candor	NC	59.72	46.61	-13.11
Coweeta	NC	46.17	36.5	-9.67
Cranberry	NC	58.35	49.7	-8.65
Beaufort	NC	45.79	34.26	-11.53
Prince Edward	VA	56.62	46.28	-10.34
Shenandoah NP	VA	63.29	54.16	-9.13
Horton Station	VA	62.1	51.56	-10.54
Sumatra	FL	41.53	33.83	-7.7
Bondville	IL	56.27	46.48	-9.79
Perkinstown	WI	43.77	40.36	-3.41
Oxford	OH	60.5	52.61	-7.89
Glacier NP	MT	35.91	40.02	4.11
Chiricahua	AZ	50.87	54.36	3.49
Grand Canyon	AZ	57.23	54.63	-2.60
Gothic	CO	54.34	50.34	-4.00
Lassen Volcanic	CA	55.15	51.8	-3.35

Table 2.5: Results from each method applied on deseasonalized temperature data from CASTNET sites. * is marked if the test statistic is lower than 95% critical values according to Table 2.2.

		Parametric		Non-parametric			
		Mean		Mean		Variability	
Site	State	Optimal c	P	Optimal c	W	Optimal c	M
Ashland	ME	1621	3.88	1621	4.12	1461	1.95*
Howland	ME	1637	3.91	1595	2.57*	677	1.89*
Woodstock	NH	1356	1.46*	1618	3.86	1496	2.38*
Connecticut Hill	NY	1189	6.91	1189	6.3	1492	1.96*
Penn State	PA	1187	6.83	1187	6.69	430	1.23*
Parsons	WV	1620	4.49	1620	4.32	442	1.61*
Edgar Evins	TN	1329	8.16	1329	7.27	1442	5.26
Georgia Station	GA	1270	10.01	1270	9.58	861	4.01
Sand Mountain	AL	1270	9.82	1270	9.64	826	3.14*
Candor	NC	1225	6.45	1225	5.86	1501	3.27*
Coweeta	NC	1446	11.24	1446	10.53	724	2.43*
Cranberry	NC	1242	8.85	1242	8.44	1428	2.61*
Beaufort	NC	1641	10.52	1641	9.7	1633	4.58
Prince Edward	VA	1200	9	1200	8.68	1625	1.98*
Shenandoah NP	VA	891	5.95	891	5.56	1050	2.55*
Horton Station	VA	1234	7.25	1242	6.87	1250	1.32*
Sumatra	FL	532	5.99	533	6.01	1438	3.89
Bondville	IL	1656	7.65	1647	6.87	560	3.26*
Perkinstown	WI	1196	8.39	1196	8.19	563	3.83
Oxford	OH	1330	6.2	1330	5.87	1440	1.91*
Glacier NP	MT	580	6.57	580	5.93	1444	3.83
Chiricahua	AZ	910	8.46	910	8.05	1247	2.27*
Grand Canyon	AZ	777	6.22	777	5.82	1057	1.52*
Gothic	CO	912	10.65	912	10.21	834	2.18*
Lassen Volcanic	CA	894	4.33	894	4.27	1096	3.41

Table 2.6: Temperature means ($^{\circ}C$) for each CASTNET sites before and after the changepoint. The differences between before and after the changepoints are also shown. The change time c is based on the non-parametric method. N/A represents the CASTNET sites that we do not detect any change.

		Temperature means ($^{\circ}C$)		
Site	State	Before c ($\bar{X}_1$)	After c ($\bar{X}_2$)	Difference ($\bar{X}_2 - \bar{X}_1$)
Ashland	ME	21.14	21.92	0.79
Howland	ME	N/A	N/A	N/A
Woodstock	NH	21.94	22.1	0.16
Connecticut Hill	NY	21.47	22.57	1.09
Penn State	PA	23.88	24.88	1
Parsons	WV	24.16	25	0.84
Edgar Evins	TN	26.7	27.82	1.12
Georgia Station	GA	27.96	29.25	1.29
Sand Mountain	AL	27.04	28.28	1.24
Candor	NC	27.76	28.8	1.04
Coweeta	NC	24.81	26.29	1.48
Cranberry	NC	20.08	21.14	1.06
Beaufort	NC	27.47	28.74	1.27
Prince Edward	VA	27.12	28.52	1.41
Shenandoah NP	VA	19.5	20.33	0.83
Horton Station	VA	21.96	22.9	0.94
Sumatra	FL	29.03	29.77	0.74
Bondville	IL	26.11	27.59	1.48
Perkinstown	WI	21.32	22.66	1.34
Oxford	OH	25.76	26.85	1.09
Glacier NP	MT	20.21	22.08	1.87
Chiricahua	AZ	27.79	28.88	1.08
Grand Canyon	AZ	23.81	24.82	1.01
Gothic	CO	15.83	17.53	1.7
Lassen Volcanic	CA	20.15	20.96	0.8

2.5 Conclusions

In this chapter, we provide two methods for detecting changepoints. The first one assumes that the data is drawn from normal distributions which is not always reasonable in practice. This method is based on the two-sample t-test by assuming that the possible break point is at c, and then compare between the means of two sub-series: the one before time c, and the other one after time c. In reality, c is unknown, so we cannot use the two-sample t-test right away. We use the maximum test statistics among the possible break points c's and estimate a critical value that can provide enough significance to decide if c is actually a changepoint of the time series. The second method avoids the assumption of normality. This is a non-parametric variant which uses ranks of the data to find changepoints. Since this method is non-parametric, it is safer to apply on data without distributional assumptions. The method is based on the Mann-Whitney test. We find the most significant changepoint by maximizing the test statistics over the time index c and argmax is the change time if the test statistic is above the critical value. We should also note that we do not consider changes near endpoints. For the data from CASTNET, we expect the changepoints of CASTNET data to be in the middle of the time series (if they exist at all), so the issue of endpoints could be neglected.

Surprisingly, most of the changepoints calculated by the parametric case match the changepoints caculated by the non-parametric method. We can use either method but as we mentioned before that it may be a safer way to stick with the non-parametric way. These results from CASTNET also show that after the NO_X SIP call was implemented in eastern U.S., the means of ozone have decreased within 2-4 years (2002-2004). However, the changes in temperature means are fluctuated among the CASTNET sites. We would also like to see more aspects of the the

policy in the long run. To do that, we move to an analysis of extreme in ozone which will be covered in the next chapter.

CHAPTER 3

TAIL BEHAVIOR OF THE SURFACE OZONE DISTRIBUTIONS

3.1 Introduction

Rieder et al., (2013)[41] suggested that the ozone distribution in the eastern U.S. has a summertime temporal distribution with heavier than Gaussian tails. It is the more extreme events that exceed the Environmental Protection Agency (EPA) ozone regulated limits. In this chapter, we use extreme value theory to analyze high ozone and temperature concentrations over the continental U.S. The analysis of the CASTNET dataset allows us to examine continental scale geographic differences in extreme behavior of ozone and how this behavior changes following the NO_X SIP implementation. CASTNET provides a regional characterization of ozone concentrations representative of rural background conditions, and is thus more suitable for an evaluation of global chemistry climate models than data ozone measurements more specific to local conditions. In contrast to Rieder et al., (2013)[41], we present results over the entire U.S. and focus on the shape of the tail of the ozone distribution. We show that at some locations the ozone probability distribution is not exponentially bounded, and thus can be characterized as heavy tailed; however, in other locations the distribution is not heavy tailed. In these latter locations the distribution is bounded and the ozone concentration has an upper limit for arbitrarily long return periods. In these locations we characterize the distribution as having light tails. We show that following the reduction of NO_X emissions in association with the NO_X SIP call the number of locations where the ozone distribution is heavy tailed increases significantly.

This chapter is organized in the following manner. In Section 3.2, we recall the

data used in Chapter 2 and how it is further transformed so as to apply extreme value theory to the CASTNET ozone measurements. In Section 3.3, we review extreme value theory and apply it to synthetic distributions so as to recommend a best practice procedure for applying extreme value theory to measured ozone distributions. Our contention is that applying extreme value theory and estimating extreme behavior is trickier than it may appear initially, for several reasons. In Section 3.4 we analyze the geography of extreme values throughout the U.S. using CASTNET data and at the end in Section 3.5 we give our conclusions.

3.2 Data

Similar to the previous chapter, we use the summertime data from CASTNET to analyze the extreme ozone. As we have seen from the previous chapter that most changepoints happen around the midpoint of time series, for simplicity, we assume that they all have the changepoint at the exact midpoint of the time series. That is, we divided the CASTNET data into 2 periods: 1992-2002 and 2003-2013 so as to investigate the effect of the NO_X SIP call on the ozone extremes. Table 3.1 summarizes the basic information from each CASTNET site divided into two periods.

In addition to the transformation in the previous chapter, we further transform the daily observations in order to make them approximately stationary. Such potential distributional changes can be thought of as arising from two different phenomena: overall inter-annual changes from year to year and intra-annual seasonal variations within the same year. We already described the intra-annual effects in Section 2.2 and how to get rid of such occurrence. For this chapter, we will add

another transformation which leads to a two-step method of data transformation. In the first step, we reduce the year-to-year variability, and in the second step we use deseasonalization described in Section 2.2. We call the first step the G-method.

G-Method (Minimizing the year-to-year variability): The idea is to estimate the average ozone concentration for each summer of each year and to subtract it from the data, while keeping the extreme values significant. Thus our calculation of the ozone extremes does not include the component due to inter-annual variability. We expect the overall ozone concentrations to decrease over time due to emission reductions and in particular due to the NO_X SIP call. While these overall changes could be corrected through a linear regression of ozone over the time period considered, the residual with respect to the regressed fit will be large due to the fact that the sample size of our ozone observations is relatively small and the inter-annual variability is large. To avoid this problem we calculate, for each summer of each year y, the average of daily observations $x_{y,d}$ of MDA8, with a certain number a of the highest extreme observations omitted. The latter precaution is taken in order to avoid letting the extremes affect the estimate of the yearly mean level of the ozone concentration. We denote the latter average by $m_{y,a}$ and define, for each year y and each day d,

$$\hat{x}^G_{y,d} = x_{y,d} - m_{y,a}\,. \tag{3.1}$$

In our analysis, we use $a = 10$ as the default value.

For the two-step transformation, we will use G-method first and then apply the deseasonalization. Therefore, the formula in (2.2) now becomes

$$\hat{x}^{DS}_{y,d} = \frac{\hat{x}^G_{y,d} - m_d}{sd_d}\,. \tag{3.2}$$

Recall that Figure 2.1 shows the smoothed means and smoothed standard deviations and we notice that if we apply G-method first and then the smoothed

standard deviation curve is not far from a constant value. This behavior is also typical for other stations. Therefore, for simplicity, we forgo dividing by the standard deviation in (3.2) at all stations and use, instead, as the standardization of the daily values,

$$\hat{x}^{DS}_{y,d} = \hat{x}^{G}_{y,d} - m_d\,. \tag{3.3}$$

In some cases, the deseasonalized data in (3.2), which are in a normalized scale, cannot tell us the actual levels of the measurements (ozone or temperature), thus, we have to transform the data back to its original scale while not breaking the stationarity of the data. We call this rescaled data.

Rescale (reverting deseasonalized data to the original scale): We revert the processes from G-method and deseasonalization by multiplying the overall mean of standard deviations and add the averages of means from seasonal cycles and year-to-year effects. In particular, suppose we have the deseasonalized ozone data $\hat{x}^{DS}_{y,d}$, we will use the following formula to rescale it back to its original scale:

$$\hat{x}^{res}_{y,d} = \hat{x}^{DS}_{y,d} \times \left(\frac{1}{D}\sum_{d'=1}^{D} sd_{d'}\right) + \frac{1}{D}\sum_{d'=1}^{D} m_{d'} + \frac{1}{Y}\sum_{y'=1}^{Y} m_{y',a}. \tag{3.4}$$

Again, we use $a = 10$ as the default value for year-to-year effects. In our study, we use the data from CASTNET only during summertime which covers only 92 days per year, so the number of days per year is $D = 92$ and the number of years is $Y = 22$. If we use the transformation in (3.3), then we can take the standard deviation terms to be 1.

Table 3.1: Basic Information for each site that we use in our analysis. Period 1 is from 1992 to 2002 and period 2 is from 2003 to 2013.

Site name	State	Period 1		Period 2	
		Ozone [ppb]	Temperature [°C]	Ozone [ppb]	Temperature [°C]
		mean (sd)	mean (sd)	mean (sd)	mean (sd)
Ashland	ME	37.79 (10.64)	21.18 (4.24)	32.77 (8.94)	21.41 (4.10)
Howland	ME	40.08 (13.07)	22.70 (4.18)	35.01 (10.46)	22.71 (4.17)
Woodstock	NH	38.66 (11.62)	21.98 (3.85)	34.10 (9.56)	21.96 (7.83)
Connecticut Hill	NY	54.88 (13.77)	21.58 (3.90)	46.87 (11.09)	22.27 (3.89)
Penn State	PA	59.01 (16.24)	24.00 (3.63)	50.34 (12.38)	24.58 (3.56)
Parsons	WV	60.50 (14.78)	25.88 (3.57)	52.61 (13.09)	26.34 (3.49)
Edgar Evins	TN	53.96 (12.88)	26.90 (2.86)	46.41 (11.47)	27.27 (3.03)
Georgia Station	GA	56.78 (19.22)	28.11 (2.90)	49.18 (15.50)	28.79 (2.72)
Sand Mountain	AL	59.38 (15.50)	27.19 (2.85)	49.89 (12.68)	27.83 (2.76)
Candor	NC	59.33 (16.06)	27.95 (3.50)	46.77 (13.48)	28.39 (3.08)
Coweeta	NC	46.12 (12.28)	24.80 (2.80)	36.67 (10.74)	25.68 (2.67)
Cranberry	NC	58.25 (11.31)	20.22 (2.64)	49.75 (9.73)	20.76 (2.54)
Beaufort	NC	48.08 (14.96)	27.31 (2.33)	41.73 (12.66)	28.09 (2.24)
Prince Edward	VA	56.34 (12.70)	24.30 (3.23)	46.70 (10.31)	24.33 (3.05)
Shenandoah NP	VA	56.51 (13.97)	27.30 (3.63)	46.35 (11.15)	28.09 (3.51)
Horton Station	VA	63.20 (12.39)	19.66 (3.25)	54.21 (9.98)	20.26 (3.02)
Sumatra	FL	61.99 (11.94)	22.11 (2.94)	51.59 (10.13)	22.56 (2.96)

Table 3.1: (Continued).

Site name	State	Period 1 Ozone [ppb] mean (sd)	Period 1 Temperature [°C] mean (sd)	Period 2 Ozone [ppb] mean (sd)	Period 2 Temperature [°C] mean (sd)
Bondville	IL	40.48 (14.18)	29.51 (2.29)	33.42 (11.49)	29.64 (2.38)
Perkinstown	WI	57.31 (13.96)	26.21 (3.70)	49.36 (12.52)	26.59 (3.52)
Oxford	OH	44.02 (12.21)	21.41 (3.92)	41.43 (10.67)	22.39 (3.89)
Glacier NP	MT	36.28 (9.19)	21.20 (5.47)	40.12 (8.57)	21.96 (5.56)
Chiricahua	AZ	53.48 (8.28)	27.91 (3.38)	54.57 (8.34)	28.87 (3.30)
Grand Canyon	AZ	56.60 (7.86)	24.15 (3.72)	56.47 (7.44)	24.70 (3.24)
Gothic	CO	54.29 (7.39)	16.16 (3.72)	52.50 (7.16)	17.40 (3.53)
Lassen Volcanic	CA	54.80 (10.61)	20.48 (4.67)	53.78 (9.99)	20.83 (4.77)

3.3 Methodology

In this section we discuss methodologies for finding extreme values in the CASTNET ozone data. Recall that the length of this dataset is not extensive, only consisting of less than 1000 data points for each of the two periods examined (1992-2002 and 2003-2013). In particular, we discuss the Maximum Likelihood Estimation procedure that assumes a parametric model, the Generalized Pareto Distribution, described below. We also consider a Hill estimator-based approach; see de Haan and Ferreira, (2006) [9]. These estimators are tested against synthetic data distributions where the extreme values are known. The synthetic data distributions are defined in Subsection 3.3.4.

3.3.1 Regular Variation

To understand the basic characteristics of extreme distributions, we should introduce the notion of *regular variation*.

A measurable function $F : \mathbb{R}^+ \to \mathbb{R}^+$ is *regularly varying* at ∞ with an index $\alpha \in \mathbb{R}$ if for all $x > 0$,

$$\lim_{t\to\infty} \frac{F(tx)}{F(t)} = x^\alpha. \tag{3.5}$$

If $\alpha = 0$, we say F is slowly varying. Therefore, we can always write a regularly varying function in the form

$$F(x) = x^\alpha L(x), \tag{3.6}$$

where $L(x)$ is a slowly varying function. The notion of regular variations can also be extended to multivariate case which will be covered later, see Section 4.3.1.

Regularly varying functions are studied in many fields and one of the applications that we will use here is to estimate the index α of extreme ozone distributions.

3.3.2 Models, tail index, upper limits, and return levels

A versatile model that can handle different shape of the distributional tails is the Generalized Pareto Distribution (GPD) (see Coles, (2001)[8]), which is characterized by three parameters: location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$, and shape parameter $\xi \in \mathbb{R}$. The GPD models have been used extensively in this setting, e.g., Rieder et al., (2015)[41]. The distribution can be described by

$$F_{\mu,\sigma,\xi}(x) = \begin{cases} 1 - \left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - e^{-(x-\mu)/\sigma} & \text{if } \xi = 0. \end{cases} \tag{3.7}$$

The range of x in this distribution is $x \geq \mu$ if $\xi \geq 0$; and $\mu \leq x \leq \mu - \sigma/\xi$ if $\xi < 0$.

The shape parameter, denoted by ξ, is the most important parameter as far as the shape of the tail is concerned, and it is also the parameter with the greatest influence on the long-term return levels. Note that, if the shape parameter is negative, then the distribution has an upper limit (equal to $\mu - \sigma/\xi$) as the return times become large. If the shape parameter is positive, then the distribution is unbounded and, moreover, has "a power tail" (i.e. the tail that decreases hyperbolically fast, and not exponentially fast). This latter observation is often formalized in the statement

$$\overline{F}(x) = 1 - F(x) = P(X > x) \propto x^{-1/\xi} \text{ as } x \to \infty \tag{3.8}$$

and the number $1/\xi > 0$ of a distribution satisfying (3.8) is called the *tail index* of that distribution. The tail index is essentially the negative inverse of the regular

variation index from (3.5). It is common to view a distribution satisfying (3.8) as having a heavy tail where the ozone probability distribution is not exponentially bounded. Physically, this means that the ozone distribution has a higher probability of reaching more extreme concentrations. Later in the text, we will use $G_{\mu,\sigma,\xi}(x)$ to denote a GPD with location parameter μ, scale parameter σ and shape parameter ξ.

With this definition, a GPD with a positive shape parameter has a heavy tail and a tail index of $1/\xi$. Of course, there are many models with a heavy tail different from the GPD. However, if the shape parameter of a GPD is non-positive, then the distribution does not have a heavy tail. We have already noted that, when the shape parameter is negative, the distribution has an "upper limit" given by

$$U_{\mu,\sigma,\xi} = \mu - \frac{\sigma}{\xi}\,. \tag{3.9}$$

Usually, when a distribution F (which may or may not be a GPD) has been fitted to data, with the occurring frequency of ϕ, one computes the N-year return level, R_N, by inverting the distribution F and setting

$$R_N = F^{-1}\left(1 - \frac{1}{\phi N}\right). \tag{3.10}$$

If a distribution has an upper limit, then the bound can be viewed as the infinite time horizon return level, R_∞. An example of the relationship between shape parameters and return levels is shown in the Figure 3.1. When the shape parameter is sufficiently negative in Figure 3.1, the upper limit exists and it is close to 20-year return level. When the shape parameter approaches 0, the upper limit grows very fast and eventually is unbounded when the shape parameter becomes positive. When the shape parameter is close to 0 the upper limit can be very large, and hence, might not be reached in reality. For positive shape parameters, return levels grow exponentially fast relative to the shape parameter.

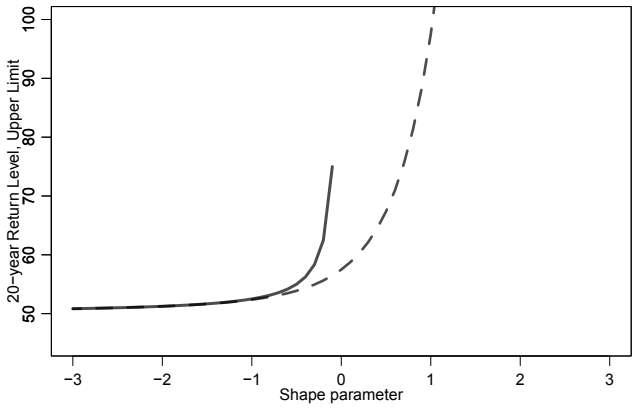


Figure 3.1: 20-year return level (dashed line) and upper limit (solid line) from $GPD_{50,2.5,\xi}(x)$ for $\xi \in [-3, 3]$.

3.3.3 Estimation

It is clear from the previous discussion that, in order to obtain reliable estimators of the return levels and, if appropriate, the upper limits, one needs to have reliable estimators of the parameters of a GPD used to fit to the data.

In Rieder et al., (2013)[41], this last step was performed using the Maximum Likelihood estimation (MLE) of the GPD parameters. Our contention is that estimating the GPD parameters is trickier than it may appear initially, for several reasons. The most important reason is that, unless one has a reason to believe that the entire data set is well described by a particular GPD, the latter model should be fitted to the part of the sample that exceeds a threshold. In Rieder et al., (2013)[41], the threshold was chosen to be 75 ppb.

It is, however, important to remember that the shape parameter ξ of a GPD describes the shape of the tail of the distribution, and, unless the threshold for

the tail has been chosen very judiciously, the MLE procedure can either attempt to find a model that fits not only the tail, but also the middle of the distribution, or alternatively, the MLE will attempt to fit the GPD to only a few data points. Finding the "right" threshold is a problem in most of the applications of GPD modeling, and it has been observed in numerous studies that the estimated shape parameter may be highly sensitive to the chosen threshold (e.g., Embrechts et al., (1997)[13]).

We propose to solve this problem by estimating the shape parameter separately, independently of the threshold chosen for fitting the GPD. Once that has been accomplished, we estimate the location parameter and the scale parameter using MLE. These parameters are not purely tail parameters, hence it is reasonable to expect that the estimated return levels are less affected by the choice of the threshold if only the location and scale parameters need to be estimated.

We suggest using the Hill estimator (see de Hann and Ferreira, (2006)[9]) to estimate the shape parameter. This estimator is a semi-parametric estimator, which is designed to estimate the shape parameter of any distribution with a power tail, i.e. satisfying (3.8). The Hill estimator, based on a sample $x_1, \ldots, x_n$, requires, as an input, a number $k = 2, \ldots, n$, and is defined by

$$H_{n,k} := \frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{x_{(n-i)}}{x_{(n-k)}} \right) , \tag{3.11}$$

where $x_{(i)}$ is the i^{th} ordered statistic of the observations $(x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)})$. We note that the Hill estimator provides an estimate of the reciprocal of the tail index, i.e., ξ. The fact that Hill estimator requires us to choose the number k of upper order statistics to use in (3.11) is a difficulty of the same nature as having to choose the threshold in a GPD model. A suggested methodology to estimate k is to plot the Hill estimator for a range of k and to look for the range where the plot

appears to be "stable"; see e.g. Resnick, (2007)[39]. This procedure is, however, sometimes difficult to implement in practice. Instead, we will use a systematic approach developed in Nguyen and Samorodnitsky, (2012)[34] to decide on the appropriate choice of k. The approach is based on sequential testing of upper order statistics. After a certain transformation, one ends up testing for exponentiality, and the test statistic suggested in Nguyen and Samorodnitsky, (2012)[34] is

$$Q_{m,n} = \frac{\sqrt{m}}{2}\left(\frac{\frac{1}{m}\sum_{i=0}^{m-1}\left(\log\frac{x_{(n-i)}}{x_{(n-m)}}\right)^2}{\left(\frac{1}{m}\sum_{i=0}^{m-1}\log\frac{x_{(n-i)}}{x_{(n-m)}}\right)^2} - 2\right). \tag{3.12}$$

One tests sequentially with increasing m, and stops to choose the value of k when the test fails:

$$k := \inf\left\{m : 1 \le m \le n, |Q_{m,n}| \ge \omega\sqrt{\frac{\theta_n}{m}}\right\} - 1, . \tag{3.13}$$

The critical value for the test statistic depends on a number $\omega > 0$ and a sequence θ_n. We follow Nguyen and Samorodnitsky, (2012)[34] and choose $\omega = 2.33$ (corresponding to the normal 99% confidence interval) and $\theta_n = (\log n)^2$.

We emphasize at this point that the Hill estimator and the procedure of choosing the number k of upper order statistics described above is designed specifically for heavy tailed distributions and, hence, would not be useful for fitting a GPD if the shape parameter is non-positive. In that case there appears to be little choice but to use the MLE. Since in application to real observations we do not know ahead of time if the best choice of estimated tail parameter is positive or not, we have implemented a multistep procedure for performing this estimation. The procedure will be described in Section 3.4.

A second important issue to keep in mind is that all the estimation approaches described above assume that the observations are identically distributed and,

preferably, independent. However, the procedures are known to be consistent even without the assumption of independence, under certain types of weak dependence. The assumption on identically distributed observations, however, requires the observations to be transformed prior to estimating the parameters as described above.

3.3.4 Synthetic data

Before estimating the extreme values in the CASTNET data, we would like to compare an estimator based on the Hill method against one based on MLE using known distributions. The idea is to generate synthetic data, in which we know the true value of the shape parameter, and apply the two competing estimation procedures to this data. We will then compare the estimates of shape parameter of the fitted GPD distribution, produced by each one of the procedures, to the true value. We will then apply the results from this analysis to an estimation of extreme values from CASTNET.

As one set of synthetic data, we have decided to generate observations from a true GPD distribution, with the shape parameters varying from -1.5 to 1.5. Rieder et al., (2013) [41] suggests that the ozone measurements can be fitted better to a GPD distribution than to a Gaussian distribution. Realizing, however, that the true ozone concentration readings may or may not be very close to a true GPD, we have also decided to generate synthetic data from a model whose tail, while possessing a well-defined shape parameter, is sufficiently modified from a close GPD fit, which we will call modified GPD. Our modification consists of using, very simply, logarithmic terms to modify a pure power tail. It is well known that such logarithmic terms do not affect the shape parameter; see Resnick, (1987) [38]

and de Haan and Ferreira, (2006)[9]. Specifically, to generate synthetic data with a positive shape parameter we use, with $\xi > 0$,

$$X = U^{-\xi}\Big(\log(1/U)\Big)^{\xi}, \tag{3.14}$$

where U is a standard uniform random variable (a random number). For such a random variable,

$$\overline{F}(x) \sim \xi^{-1} x^{-1/\xi} \log(x) \quad \text{as} \quad x \to \infty\,. \tag{3.15}$$

In the case of a negative shape parameter $\xi < 0$, the shape parameter determines how likely the observation is to be close to its ultimate upper limit. To achieve a corresponding logarithmic modification of that probability, we generate a GPD random variable Y with a negative shape parameter, and the ultimate upper limit, say, m. Then we take, as an observation,

$$X = m - (m - Y)\,|\log(m - Y)|^{\theta}\,. \tag{3.16}$$

We use $\theta = 2$. Similar to $G_{\mu,\sigma,\xi}(x)$, we will refer to these modified GPDs by denoting them as $G'_{\mu,\sigma,\xi}(x)$.

In order to create the synthetic data as close to the CASTNET data as possible, we generate 500 data sets for each model, each with 1000 data points which is approximately the same number of observations as the 11 years of summertime daily observations we have for each period. In each case, we take the location and scale parameters of the GPD to be 50 and 2.5, respectively. These parameters provide a reasonable fit to the measured ozone distribution. For each data set drawn from the synthetic distribution generated either from the true GPD model, or the logarithmically modified GPD model, we apply both the straight MLE estimate using the top 5% of the data and Hill estimator to the synthetic data (see Figures 3.2-3.4).

Table 3.2: Upper limits and return levels from period 1 (1992-2002) and period 2 (2003-2013) of ozone for each CASTNET station. For the Maximum Likelihood estimator we use data without any transformations (i.e., non-stationary, without using Equation 3.3); for Hill's, the transformations are applied to make the data stationary before using the Hill's method for distributions with heavy tails (see Section 3.4.2). Upper limits are calculated only for stations with negative shape parameters. The numbers with * are stations that have shape parameters between -0.3 and 0.3, and by our procedure described in Section 3.4.2, we set the shape parameter to 0.

		Period 1				Period 2			
		20-year Level	Return	Upper Limit		20-year Level	Return	Upper Limit	
Station	State	MLE	Hill's	MLE	Hill's	MLE	Hill's	MLE	Hill's
Ashland	ME	86.89	107.37*	104.14	-	67.89	65.71	75.46	69.77
Howland	ME	95.32	164.96	105.87	-	79.81	123.14	102.82	-
Woodstock	NH	88.70	150.92	103.27	-	69.88	112.69	76.50	-
Connecticut Hill	NY	100.78	137.85	117.21	-	100.78	142.62	209.31	-
Penn State	PA	106.91	108.60	109.41	112.25	89.64	141.01	94.23	-
Parsons	WV	99.99	150.66	117.12	-	86.43	136.85	228.68	-
Edgar Evins	TN	100.67	147.99	129.70	-	88.49	124.74	-	-
Georgia Station	GA	124.05	137.41*	150.40	-	103.86	107.65*	153.71	-
Sand Mountain	AL	100.81	100.19	101.96	101.86	91.66	107.45*	118.33	-
Candor	NC	110.00	109.95	115.05	114.72	87.73	87.39	92.21	94.83
Coweeta	NC	91.10	128.12	190.41	-	68.38	65.27	71.19	67.21
Cranberry	NC	98.53	98.99	106.46	107.70	91.65	112.40	-	-
Beaufort	NC	94.36	91.16	120.28	99.88	81.86	137.02	96.83	-
Prince Edward	VA	97.76	97.17	103.11	101.12	80.19	80.06	83.88	84.28
Shenandoah NP	VA	106.10	150.22	117.10	-	100.77	131.27	-	-
Horton Station	VA	100.81	100.72	103.90	104.48	91.48	117.46	176.14	-
Sumatra	FL	84.32	83.26	89.43	88.42	71.32	88.94*	76.19	-

Table 3.2: (Continued).

		Period 1				Period 2			
		20-year Return Level		Upper Limit		20-year Return Level		Upper Limit	
Station	State	MLE	Hill's	MLE	Hill's	MLE	Hill's	MLE	Hill's
Bondville	IL	103.41	99.44	110.01	102.91	86.96	85.87	92.14	90.64
Perkinstown	WI	93.43	132.26	132.59	-	78.49	112.00	94.88	-
Oxford	OH	110.25	171.67	117.21	-	103.00	154.35	131.16	-
Glacier NP	MT	59.16	58.22	61.90	59.28	67.11	84.04	108.50	-
Chiricahua	AZ	77.82	74.03	84.19	74.88	76.43	77.12	77.20	79.27
Grand Canyon	AZ	81.49	88.81*	90.11	-	107.96	103.24	125.02	-
Gothic	CO	76.26	76.79	77.51	80.36	78.38	105.34	96.33	-
Lassen Volcanic	CA	88.91	124.19	97.47	-	94.78	141.73	-	-

3.4 Results

In this chapter, we still use the same data from the previous chapter. However, instead of using changepoints we got from the results of the previous chapter, we assume that the changepoints of each CASTNET site is in the middle of the time series. That is, we divided the CASTNET data into 2 periods: 1992-2002 and 2003-2013 so as to investigate the effect of the NO_X SIP call on the ozone extremes.

3.4.1 Analysis of synthetic data

The MLE underestimates the true shape for positive shape parameters (Figures 3.2 and 3.3). This is true for both the true GPD distribution and the modified GDP distribution. Using the estimator based on Hill gives estimates of the shape parameter that are both less biased and less variable than the straight MLE approach. Note that both the MLE approach and the method based on Hill estimator show considerable spread in their results due to the finite number of data points used (i.e., 1000).

When the true shape parameter is negative, the estimator based on Hill estimates a shape parameter close to zero (Figure 3.4). As noted above, the Hill estimator is designed to estimate positive values of a shape parameter. On the other hand, the straight MLE approach works reasonably well for the synthetic data generated from a true GPD, although there is a large spread in the estimated shapes and the true shape parameter is somewhat underestimated (see Figure 3.4 and Table 3.3). The MLE method does not work that well for the data from a logarithmically modified model as the estimated shape parameter is concentrated

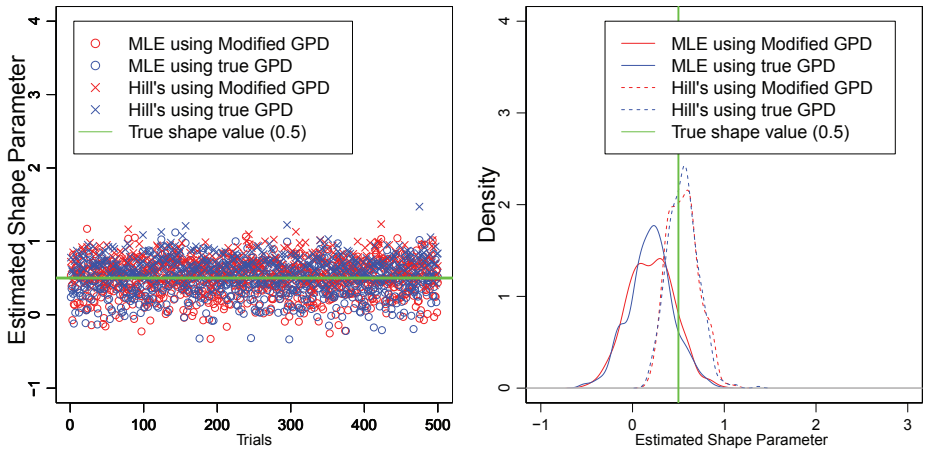


Figure 3.2: Left: Scatter plot of estimations from different methods and different models for each of 500 trials. Right: Histogram of the estimators. The data are sampled from $G_{50,2.5,0.5}(x)$ and $G'_{50,2.5,0.5}(x)$.

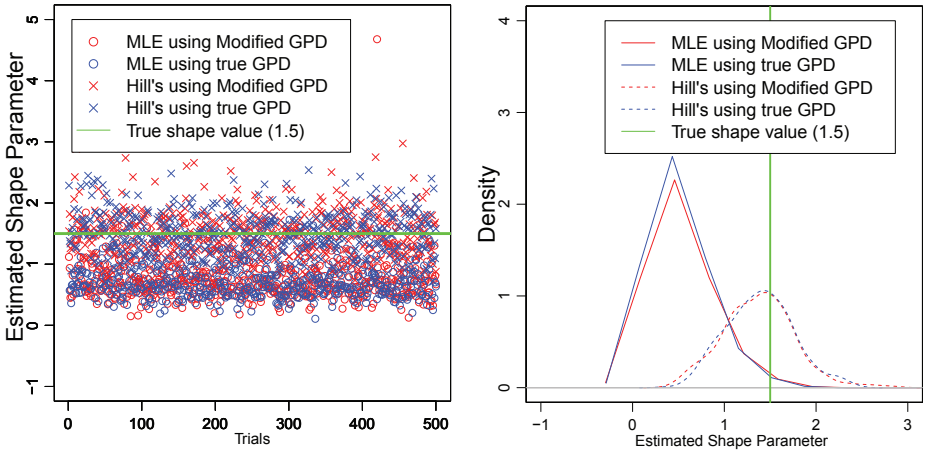


Figure 3.3: As in Figure 3.2 but for data sampled from $G_{50,2.5,1.5}(x)$ and $G'_{50,2.5,1.5}(x)$.

around the value of, approximately, -1 (see Figure 3.4).

Table 3.3 shows a summary of experiments for different models and different shape parameters for data consisting of 1000 points. The numbers in parentheses are median absolute deviations (MAD) which is defined by

$$MAD(x) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \text{median}(x)|. \tag{3.17}$$

The methodology based on Hill's method only estimates positive shapes regardless

of the true shape parameter. In particular when the true shape parameter is less than zero, the Hill's method estimates a shape parameter near zero. This methodology is quite ac- curate for shape parameters greater than 0.5. The MLE method produces fairly accurate estimates of the shape of the GPD distribution when the shape parameter is negative. It tends to estimate a positive shape parameter when the synthetic shape parameter is positive but underestimates the true shape.

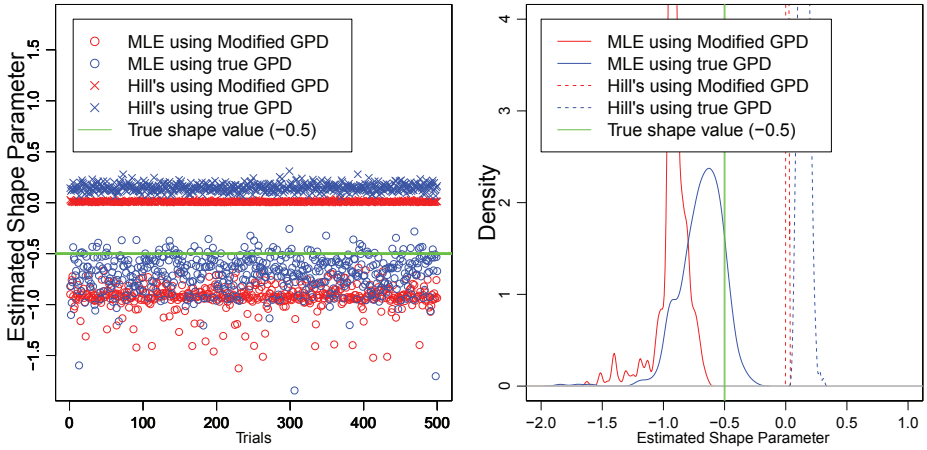


Figure 3.4: As in Figure 3.2 but for data sampled from $G_{50,2.5,-0.5}(x)$ and $G'_{50,2.5,-0.5}(x)$.

Figures 3.5-3.6 show the effect of the choice of the tail threshold in estimating the shape parameter when using MLE on a modified GPD distribution with the shape parameter of 0.5 and -0.5 (see Equations 3.14 and 3.14) in Section 3.3.4 for the description of modified GPDs). A true GPD distribution is insensitive to the choice of the tail threshold. We ran 500 trials with a sample of 1000 points in each trial and estimated the shape parameter for thresholds between the 75^{th} and 99^{th} percentile of the data.

For positive shape parameters, estimates using the top 5-10% of the data give more accurate estimates than those using the top 25% of the data; however in all cases the true shape is underestimated in the median. Using the top 1% of the

Table 3.3: Medians of estimated shape parameters taken from GPD ($G_{50,2.5,\xi}$) and modified GPD ($G'_{50,2.5,\xi}$) distributions using the MLE methodology and the Hill's methodology (see Section 3.3.3). The numbers in parentheses are median absolute deviation (MAD).

True shape	MLE		Hill's	
(ξ)	$G_{50,2.5,\xi}(x)$	$G'_{50,2.5,\xi}(x)$	$G_{50,2.5,\xi}(x)$	$G'_{50,2.5,\xi}(x)$
-1.5	-0.78 (0.06)	-0.92 (0.00)	0.02 (0.01)	0.01 (0.00)
-1.0	-0.82 (0.08)	-1.01 (0.00)	0.06 (0.02)	0.01 (0.00)
-0.8	-0.85 (0.07)	-1.08 (0.05)	0.08 (0.02)	0.01 (0.00)
-0.5	-0.69 (0.14)	-0.93 (0.14)	0.15 (0.03)	0.01 (0.00)
-0.3	-0.45 (0.14)	-1.04 (0.27)	0.21 (0.04)	0.02 (0.02)
-0.1	-0.22 (0.14)	-0.89 (0.20)	0.29 (0.05)	0.14 (0.05)
0.1	-0.02 (0.16)	-0.03 (0.15)	0.39 (0.08)	0.40 (0.08)
0.3	0.22 (0.17)	0.21 (0.18)	0.52 (0.10)	0.52 (0.10)
0.5	0.40 (0.20)	0.41 (0.20)	0.65 (0.14)	0.66 (0.14)
0.8	0.73 (0.22)	0.75 (0.22)	0.88 (0.18)	0.87 (0.18)
1.0	0.78 (0.31)	0.88 (0.57)	1.04 (0.21)	1.03 (0.22)
1.5	0.65 (4.88)	0.67 (6.23)	1.46 (0.30)	1.46 (0.30)

data gives the least accurate estimate as the estimate is subject to a severe data limitation (i.e., 10 points). Note that using the top 1% of the data gives some clustering of the results around -1. The estimation variance increases when the threshold increases, since fewer observations are used for the estimation. For negative shape parameters neither MLE nor Hill's estimator is particularly accurate. MLE shows a clustering of estimates near -1.0 consistent with the results given in right panel of Figure 3.4 for a modified GPD.

Figure 3.7 gives the return intervals as calculated from samples of 1000 data points taken from three synthetic distributions: one with a light tail ($\xi = -0.5$). one with a heavy tail ($\xi = 0.5$), and one with a shape parameter of zero. The samples are typical of the results for the synthetic distribution as shown above. In each case, we calculate the fit to the sampled distribution using both Hill's method and the MLE method. The return interval as calculated from the data is taken from the ordered statistics of the data points. Thus the highest data point

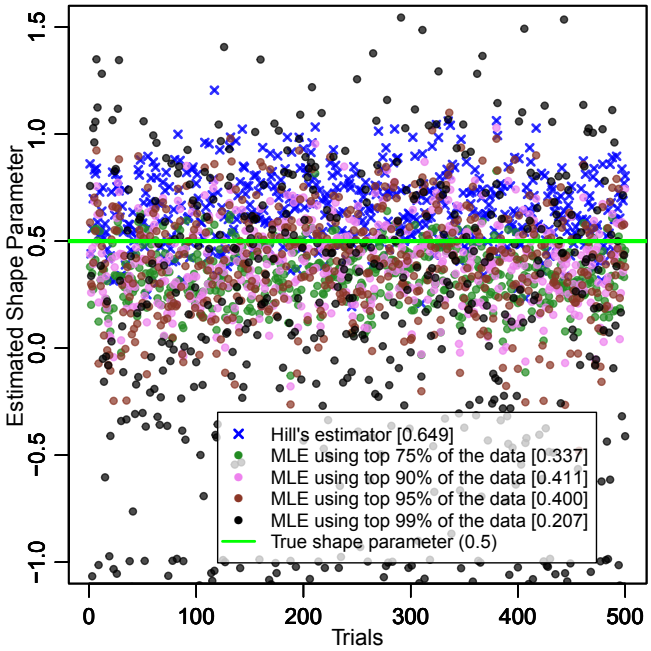


Figure 3.5: Estimated shape parameters from $G'_{50,2.5,0.5}(x)$. The experiments are run on 500 trials, with 1000 data points sampled from each trial. The numbers in brackets are medians for each case. The green line gives true shape.

is assigned a return level of the number of years represented by the sampled data assuming 100 points represent a year (recall that we are assuming the samples are only drawn for summertime data). In each case, the MLE provides the best fit to the sampled data. However, with only 1000 data points the ordered sample data tend to underestimate the true return levels (which we refer as control in the Figure 3.7) when the shape is positive. For a positive shape of 0.5, Figure 3.7 shows that Hill's method provides a better estimate of the true shape than the MLE method. MLE method tends to underestimate the true shape because it tries to fit the estimate to the observed points when most of the points do not represent the tail of the distribution. However, since the Hill's estimator cannot give negative values, the estimated return levels from Hill's estimator (blue line) deviate from the true return levels (green line) for the samples where the shape of the synthetic distribution is 0 or negative.

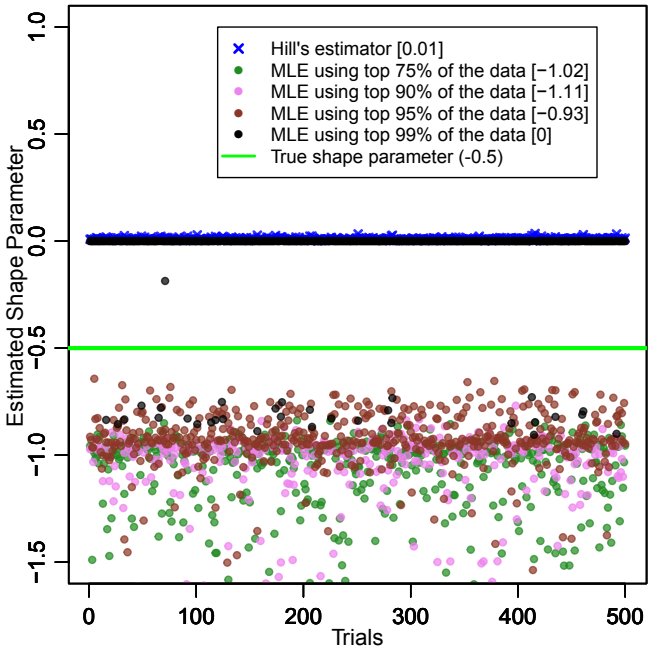


Figure 3.6: Same as Figure 3.5 but with a shape parameter -0.5 (i.e., $G'_{50,2.5,-0.5}(x)$).

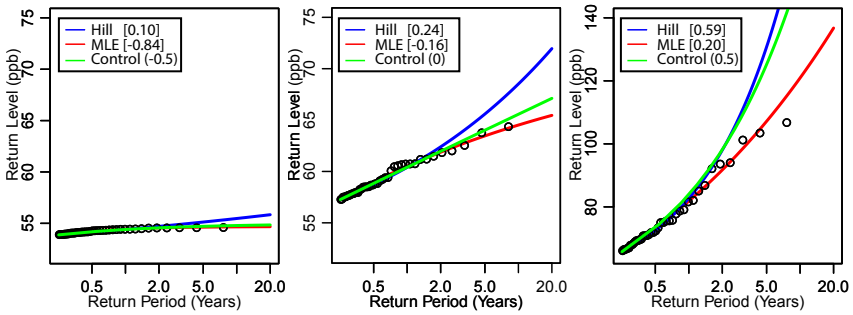


Figure 3.7: Return levels using MLE and Hill's estimators on data generated from the true distributions: $G_{50,2.5,-0.5}(x)$ (left), $G_{50,2.5,0}(x)$ (middle), $G_{50,2.5,0.5}(x)$ (right). The numbers in brackets are estimated shapes from the corresponding estimators.

3.4.2 Recommended Procedure

The synthetic data analyzed above suggests there is no ideal procedure to analyze extreme values with approximately 1000 data points. The results suggest that for distributions with a positive shape parameter Hill's methodology provides the best

fit; for distributions with a negative shape parameter the MLE method is preferred. Therefore, in our analysis of the measured ozone distributions we have devised a strategy based on the strength of both the MLE method and Hill's method (Figure 3.8).

We observe from Table 3.3 that, when the true shape parameter is non-positive, the Hill based estimator approach tends to give a small value of the shape parameter, typically below 0.3. Therefore, our approach first uses Hill's estimator. If the resulting value is above the threshold of 0.3, then this value is taken as our estimate of the shape parameter. We note from Table 3.3 that for shape parameters greater than approximately 0.3, the methodology based on Hill's estimator becomes reasonably accurate. If the estimated shape is less than 0.3 we use the MLE method to analyze the data. If the resulting shape parameter is less than -0.3, then this value of the shape parameter is taken as our best estimate of the true shape parameter. For ozone distributions where the Hill's parameter gives a shape parameter less than 0.3 and the MLE method gives a parameter great than -0.3, we are not confident that the shape parameter is significantly different from zero. Thus, by default we assign a default value of 0 to the shape parameter. Note that this value is, indeed, special, for it corresponds to a model with an unbounded range, but without heavy tails. The formal description of our procedure is given here: (also see Figure 3.8).

- Estimate the shape parameter by using the Hill estimator-based approach. Call the result α.
- If $\alpha > 0.3$, we use α as the estimate of the shape parameter.
- If $\alpha \leq 0.3$, we estimate the shape parameter by using MLE. Call the result β.

- If $\beta < -0.3$, we use β as the estimate of the shape parameter.
- If $\beta \geq -0.3$, we estimate the shape parameter as zero.

When the MLE method is used there is a choice as to the appropriate threshold to choose: if the threshold is too high the results will not be robust, if the threshold is too low the MLE methodology is more likely to underestimate the shape parameter. Our results suggest that the best choice for a threshold is the upper 5% of the data points. This gives a fixed sample size at all the stations. A fixed threshold to estimate the tail parameter is likely not to be as precise.

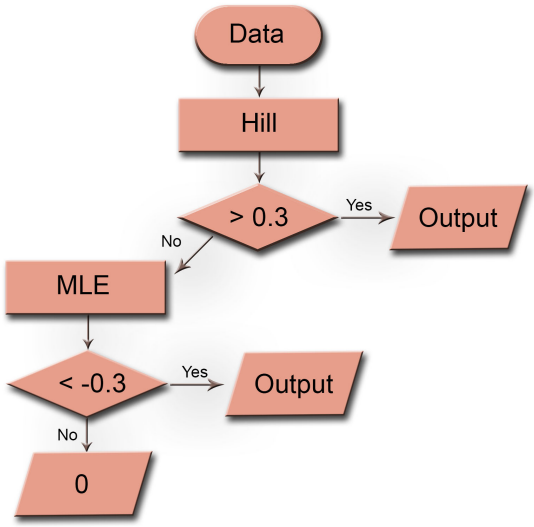


Figure 3.8: Flow chart of the procedure adapted for this study (see Section 3.4.2)

3.4.3 Application to CASTNET data

In this section we apply the methodology outlined above to the CASTNET ozone data. In particular, we are interested in better understanding the tail behavior of the measured ozone distributions. While not explicitly pointed out in Rieder et al., (2015)[40], their results would seem to suggest that the ozone distributions are

in fact bounded by an upper limit or a maximum value that ozone will not exceed (even as the return period approaches infinity). Thus from these results we would conclude that ozone behaves as a light tailed distribution. However, our analysis of synthetic ozone distributions suggest that MLE tends to underestimate the tail behavior of distributions particularly when the distributions have heavier tails. Distributions with positive tail parameters would in fact suggest that there is no readily defined upper limit to the ozone distribution. While physical limits exist, of course, heavier tail distributions would suggest more extreme behavior. Based on our analysis of synthetic distributions here, we analyze the extent to which the measured ozone distributions, in fact, exhibit heavy or light tails. In particular, we analyze the geography of the ozone extremes and the impact of the NO_X SIP call on these extremes. For this purpose, as discussed above, we divide the ozone concentration data from CASTNET into two periods (1992-2002 and 2003-2013).

Results for the return value and upper limit are given in Table 3.2. In Table 3.2, in each case we calculate twenty-year return level and the upper bound of the ozone distribution using the MLE method with no data transformations (e.g., without correcting for the seasonality or the inter-annual variability). This is similar to the methodology used in Rieder et al., (2013)[41] except we define the upper tail using 5% of the data instead of the 75 ppb cutoff. Consistent with Rieder et al., (2013)[41] the 20-year return level decreases between period 1 and period 2 for all stations located in the eastern part of the U.S., although the 20-year return period in fact increases for a number of stations located in the western U.S. between these periods. Note that for all stations in period 1 and for all but three stations in period 2 the method solely based on MLE has an upper limit, suggesting light tailed distributions. Depending on the station the upper limits generally range between 100 and 150 ppb in the eastern part of the U.S. (although Coweeta N.C.

has an upper limit of 190) and the Western Stations have an upper limit below 100 ppb. Note however, that for only approximately half of the stations analyzed in the eastern part of the country does the upper limit decrease between the two periods. These results already suggest that the behavior of the upper limit of the ozone distributions behave rather differently than the return level. We return to this point below.

In those cases where the analysis specified in Section 3.4.2 returns an upper limit, the upper limit is determined using the MLE method on the transformed ozone data (e.g., during period 1 for Penn State, and Mt, Candor etc.). As evident from these results, removing the seasonality and inter-annual variability from the data (see Section 3.2) does not dramatically change either the return levels or the upper limits. During both periods there are a number of points where the application of the Hill's method suggests the shape parameter is positive and greater than 0.3. As discussed above we suggest these stations have heavy tailed ozone distributions and do not have an upper bound. At these stations, as indicated in Table 3.2, the MLE methodology produces a notable underestimate of the large 20-year return period.

Figure 3.9 gives the estimated shapes by using the MLE estimator with different thresholds (fixed 75 ppb or the top 5% of the data) and Hill's estimator from period 1 and period 2. The brown lines divide stations into four regions: northeastern U.S., southeastern U.S., Midwest, and western U.S. Results with a fixed threshold of 75 ppb (as used in Rieder et al., (2013)[41]) tend to underestimate the shape parameter compared with MLE using a threshold of 95% or using the Hill's method. Consistent with the results above, Hill's estimator tends to estimate a larger shape parameter than the MLE with a cutoff of 75 ppb and there is some clustering about

-1 for the estimated shape parameters using MLE. Clustering around -1 was noted when using MLE to analyze synthetic data distributions (Figures 3.5-3.6)

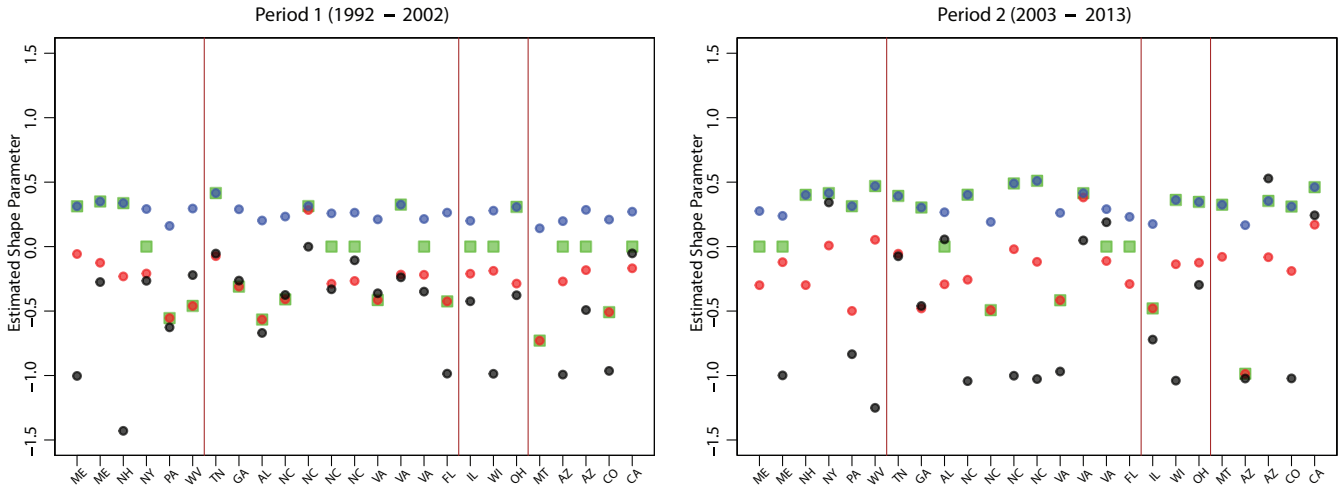


Figure 3.9: Estimated shapes from period 1 [1992-2002] (left) and period 2 [2003-2013] (right) using Hill's estimator and different thresholds for the MLE estimator for each of the stations given in Table 3.2, grouped by regions. Black: MLE method using 75 ppb as the threshold. Red: MLE method using the top 5% of the data as the threshold. Blue: Hill method. Green: Our procedure described in Section 3.4.2.

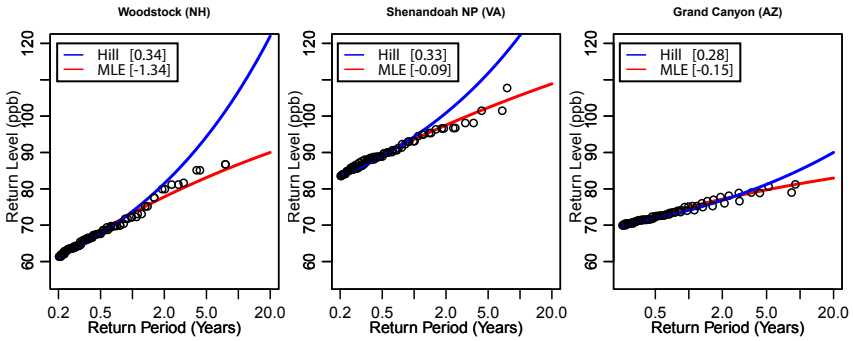


Figure 3.10: Estimated return levels using MLE and Hill's estimators at selected stations. Numbers in brackets are the estimated shape from the corresponding estimators.

Ozone data fitted by using the MLE method and the Hill's method is given in Figure 3.10 for three selected stations. Note the similarity between these fits and those shown for the synthetic data (Figure 3.7). Our analysis procedure suggests that the ozone at Woodstock station (left) and Shenandoah National Park station (middle) have positive shape parameters. As in the synthetic data, the Hill estimator deviates from the data points for the larger return levels. The shape parameter at Grand Canyon (right) tends to be close to zero as neither Hill's nor MLE estimators give a robust result.

The resulting estimates of the shape parameters, return levels, and, if applicable, ultimate upper limits across selected stations are summarized graphically in Figure 3.11 for the two periods. Stations with upper limits (negative shapes) occur in all parts of the U.S. during both periods with no easily identifiable regional clustering. In agreement with the results of Rieder et al., (2013) [41] and indicated above in reference to Table 3.2, we find the large 20-year return periods for sites in period 2 decrease with respect to period 1 throughout the Eastern U.S. Note in each period, the general correspondence between return period and whether the distributions have an upper limit using our method. Distributions without an

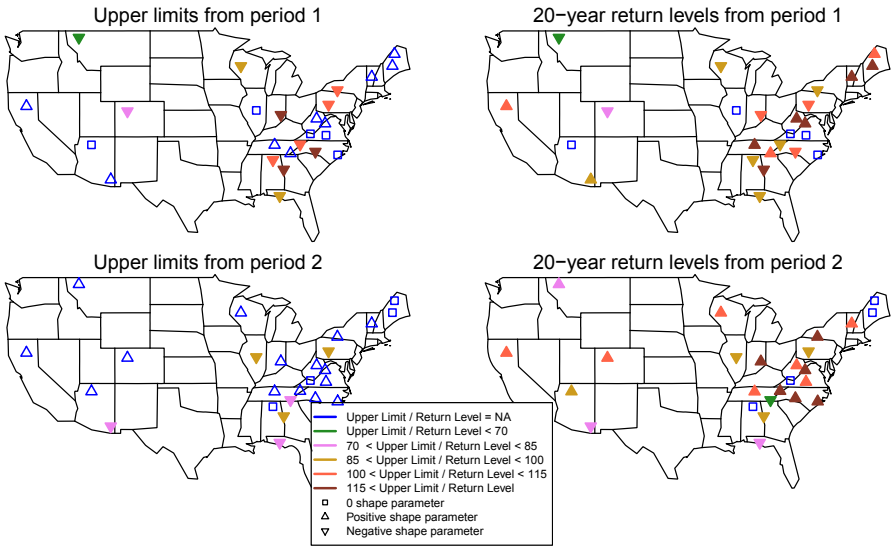


Figure 3.11: Estimated upper limits and return levels for period 1 [1992-2002] (left) and period 2 [2003-2013] (right) using the procedure in Section 3.4.2. Maps on the left side represent upper limits; hence, we label upper limits as NA for the sites with non-negative shape parameters. Similarly, maps on the right sides represent 20-year return levels; NAs are labeled when the shape parameter is 0

upper limit generally have larger return periods than those that have an upper limit.

Significant differences between the number of distributions with an upper limit are noted between the two periods. These differences are summarized in Figure 3.12. This figure suggests that the shape parameters have overall become more positive between period 1 and period 2. We suggest that this difference can most likely be attributed to the NO_X SIP call. The fact that a number of stations have larger upper limits during period 2 than period 1 is also evident, simply using an MLE analysis (see Table 3.2). In the Western States the shape parameter has also become more positive although this cannot be attributed to the NO_X SIP call.

We can use a binomial model to quickly check whether the difference in shape between the two periods is statistically significant. Assume that the probability

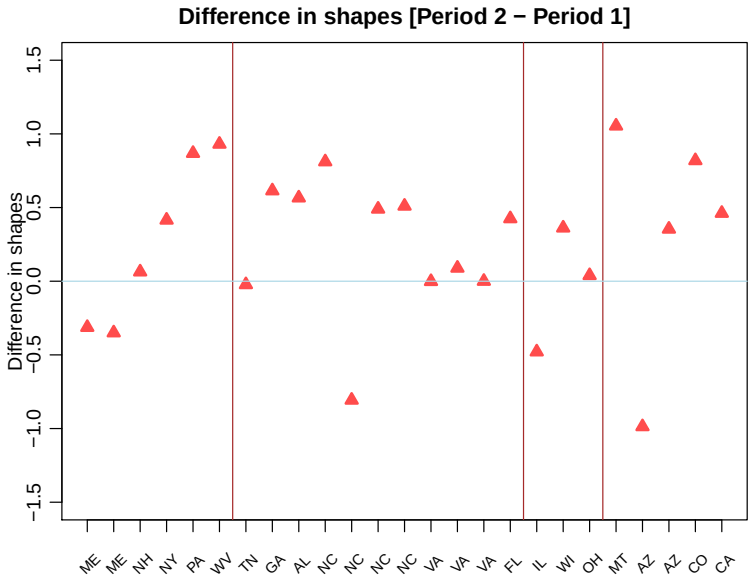


Figure 3.12: Difference between the estimated shapes between period 2 (2003-2013) and period 1 (1992-2002) for each station using the procedure adapted in Section 3.4.2.

of changing sign from period 1 to period 2 is $p = 0.5$. Let X be the number of stations where the shape parameter decreases from period 1 to period 2. According to Figure 3.12, we have 8 stations that have decreasing shapes from period 1 to period 2. This gives $P(X \leq 8) \approx 0.054$. Therefore, it is unlikely that the probability $p = 0.5$. This suggests that the increase in shape from period 1 to period 2 is statistically significant.

3.5 Conclusions

In this chapter, we have explored estimates of the tails of the ozone distribution from CASTNET data within the continental U.S. Whether the tails of the distribution are light or heavy is critical in determining the return period of the distribution and whether the distribution has an upper limit. We contend that determining

the tail parameter of ozone distributions is trickier than it may appear.

The Hill's estimator approach is designed to work on distributions with a heavy tail, especially when the tail parameters $\xi > 0.3$ this method is more accurate than the MLE method which tends to underestimate the tail parameter. However, for distributions with a negative tail parameter (light tailed distributions), our only resort is to use the MLE method. However, the MLE method is sensitive to the number of points included in the tail of the distribution. From Section 3.4, we find that fitting approximately the upper 5% of the dataset to the GPD distribution results in the best answer. According to the results from Section 3.4, we have proposed an analysis methodology whereby ozone data from CASTNET is analyzed using a procedure that utilizes the relative strengths of both the MLE and Hill's methods.

Previous studies have found that pollutant distributions are approximately log-normal (e.g., Ott, (1990)[36]). A log-normal distribution has a shape parameter of zero and can be classified as neither light nor heavy tailed although it is unbounded. For the CASTNET stations, the estimated distribution shape of the ozone distribution as analyzed either using the MLE methodology or the Hill's methodology generally ranges from -0.5 to +0.5. This is in the critical range about a shape of zero. For shapes between -0.5 and 0.0, the upper bound is also sensitively dependent on the shape parameter. For shapes between 0 and 1, the return interval is heavily dependent on the shape parameter.

Analysis of synthetic distributions in Section 3.4 gives considerable spread in tail estimates. However, our analysis of the synthetic data using the two methodologies gives us confidence that in certain locations the measured ozone distribution is in fact heavy tailed although in other locations, the distribution is light-tailed.

When analyzing the 5-year return interval the distinction between heavy tailed and light tailed (see Figure 3.7) is not critically important, with return level difference within 10 ppb. However, in an analysis of more heavy tailed distributions, this distinction is of critical importance in understanding the extreme behavior. It is also important in considering whether a distribution has an upper bound.

The return interval is dependent on the location, scale and shape of the distribution. Consistent with Rieder et al., (2013)[41], we find that the location of the ozone distribution is dependent on emissions, decreasing as emissions decrease. Thus with no change in scale or shape, we would expect the ozone concentrations for a given return interval to decrease as emissions decrease. In fact, Rieder et al., (2013)[41] finds that the 5-year ozone return levels decrease over the eastern portion of the U.S. as emissions are reduced with the SIP call. Here, however, we find strong evidence that on average the tails become heavier following the NO_X SIP call as the shape parameter increases at the majority of the sites. We find that the upper limits of only three stations in the eastern U.S. decrease between periods. In fact, we find that in many locations the distribution goes from light tailed to heavy tailed following the SIP call. For 20-year return levels our methodology suggests that the ozone concentration increases at 7 out of the 20 stations over the eastern U.S. following the SIP call. This stands in contrast, but is not necessarily inconsistent with Rieder et al, (2015)[40] who finds that the 90 percentile of the ozone distribution decreases faster than the mode as future NO_X emissions decrease. We hypothesize, without proof that higher NO_X emissions may act to better buffer against very high ozone concentrations, acting to decrease the tails of the distributions.

The fact that ozone distributions may have an upper bound dependent on

location and emissions has not been well explored in the literature, but likely has implications for pollution policy. The upper bound is an additional measure of extreme ozone behavior, addressing the question of how bad it can get. These distributions have less extreme behavior than those that are unbounded. The upper limit is an extrapolation from available statistics and in fact might not ever be reached in reality. In some locations we find that an upper bound does not exist and thus the distribution can be characterized as heavy tailed characterized by more extreme behavior. While emission reductions may reduce the ozone concentrations for short return intervals, they may, in fact, result in a heavier tailed distribution.

CHAPTER 4

MULTIVARIATE CASE: EXTREMAL DEPENDENCE BETWEEN OZONE AND TEMPERATURE

4.1 Introduction

In this chapter, we are moving away from the study in effects of the NO_X SIP call imposed in early 2000s. Instead, we would like to study the relationship between high level of temperature and ozone. Usually, the studies of temperature and ozone do not concentrate on extreme distributions and the relationship between surface ozone and temperature was found to be linear (Steiner et al., (2010)[47]). However, when it comes to extreme values, the relationship between temperature and ozone becomes more complicated for many reasons suggested in Shen et al., (2016)[44]. When we consider only the high observations, the distributions of ozone and temperature become highly non-Gaussian and the relationship becomes non-linear (Wilson et al., (2014)[54]). This phenomena makes linear regressions unlikely to correctly predict the extremes. Shen et al., (2016)[44] also suggests that when the temperature is high enough, an ozone suppression appears. This makes it necessary for us to look carefully at both temperature and ozone tail distributions and their non-linear relationships.

In this chapter, we look at three different models. Two of them represent the present climate (around 1992 to 2025, depending on the model–we will explain this in details later) and the other one represents possible future climate, predicted to correspond to 2100 - 2125. The goal is to evaluate the capability of the models to capture the extreme behavior of temperature and ozone. To do this, we need to understand the relationship between extreme temperature and ozone,

and quantify it. In practice, we do not want to use correlations to measure the dependency between extreme temperature and ozone because correlation does not give information about extremes. The standard way to compare extreme values is when they have regular variations of the same tail index (see 3.3.1 for univariate regular variations and 4.3.1 for multivariate regular variations). In this chapter, we develop a procedure that transforms the data and quantifies the relationship between extreme temperature and ozone.

This program is pursued in this chapter which is organized as follows: in Section 4.2, we describe the data—how the data are obtained, how the data are organized, and how the data are processed before using it in our study; in Section 4.3, we introduce the multivariate regular variations, revisit Hill's estimator, and also describe how we measure and quantify the relationship of extreme variables; in Section 4.4, we apply conventional methodologies, such as calculating means, standard deviations and correlations, and compare the results from observations and models. Then, we apply our methodology to the extreme temperature and ozone generated by the models as well as to the data from the observations. Based on our analysis, we discuss how well these models can represent the extreme temperature and ozone.

4.2 Data and Model Descriptions

In addition to CASTNET data we used in previous chapters, we would like to determine if the models are adequate for predicting future changes, so we add 3 different simulations obtained from the Community Earth System Model (CESM: www.cesm.ucar.edu) and Computational and Information Systems Lab (CISL:

www.cisl.ucar.edu). The first model is a simulation that uses observed meteorological data to assimilate the global climate and the other two models are general circulation models (GCMs). We selected these models because they are used frequently in climate studies. The details of each model will be explained below. For the comparison purpose, we use CASTNET data covering the summertime of 1992-2011 to match the data length from two of the models.

For the first model, we will call it *specified dynamics model*, or *SDM* in short. This model assimilates the global climate with observable physical inputs, for example, wind component, surface temperature, surface pressure, heat flux, etc. The model includes the chemical emissions of ozone precursors. The chemical reactions are solved internally within the model using numerical solvers (Lamarque et al., (2012)[27]). We use the model with specified sea-surface, sea-ice distributions and meteorological fields observed from real data. However, we should note that the model takes specified data at climate levels (i.e., means and standard deviations). This model can capture the ozone in troposphere and stratosphere with the grid resolution of $1.9^\circ \times 2.5^\circ$ with 26 vertical levels, but we only consider the surface level of ozone. We also use temperature, which is an input for the model, as a covariate of surface ozone in our analysis. See the full detailed analysis of this model in Lamarque et al., (2012)[27] which talks about specific parameters, the equations behind the simulations and the biases from the simulations. For our setup, we use surface temperature and ozone between 1992-2010 (19 years) from this model in our analysis.

The other two models are from general circulation models (GCMs) which use a mathematical model to predict the atmosphere and oceans circulations. They are different from the specified dynamics model in that they calculate meteorological

fields internally instead of taking them as inputs from observations. GCMs are widely used in weather forecast and climate studies. The benefit of using GCMs is that we can obtain longer time series than the specified dynamics model and they can also be used to examine future conditions; however, the drawback is that they might not be consistent with the data from real observations for specific years but the average of the entire series should be consistent to some extent. We use the data from this model from two different time frames. One is from 2000-2025 and the other one is from 2100-2125.The difference between these two models is that the model representing future climate (2100-2125) has higher average ozone and average temperature than the model representing current climate (2000-2025). In this analysis, we truncate the first 6 years of each period because the simulations need some time to stabilize and avoid biases that could occur from the initial conditions. Therefore, we only use the data from 2006-2025 and 2106-2125 which will be called GCM 2000 and GCM 2100, respectively. Similar to the specified dynamics model, these GCMs have the grid resolution of $1.9^\circ \times 2.5^\circ$ as well.

These three models generate ozone and temperature (and much more meteorological data) over the globe; however, we reduce our scope down to continental U.S. only, so we crop out the data to covers from about 230°E - 300°E and 21°N - 51°N. The setup looks very similar to the previous chapter. For each year, we will use only the summertime data (June, July and August–JJA). For ozone, we still use the maximum daily 8-hour average (MDA8), while we use daily maximum for the temperature. Both data have been processed using G-method (3.1) and deseasonalization (3.3) described in the previous chapter to get rid of the year-to-year effects and seasonal cycle effects.

4.3 Methodology

In this chapter, we use conventional methods such as calculating means and correlations to examine general aspects of the data. We also consider less conventional methods which will be explained in this section. In addition to the previous chapters, we take temperature as a covariate of ozone for our analysis. Generally, temperature and ozone are correlated but the dependence is more complicated in the case of extremes. We would like to determine the relationship between the extremes of temperature and ozone. The study is based on the regular variation framework that was described in Section 3.3.1. The section 4.3.1 extends the concept to multivariate case.

4.3.1 Multivariate Regular Variation

Let $\mathbf{X} := (X_1, ..., X_n) \in \mathbb{R}^d$ be a non-negative random vector with a distribution F. We say F is *regularly varying* if there exists a measure on $\mathbb{R}^d$ that is finite on sets bounded away from the origin such that

$$\lim_{t\to\infty} \frac{\Pr(\mathbf{X} \in tB)}{\Pr(|\mathbf{X}| > t)} = \mu(B), \tag{4.1}$$

where B is a Borel set bounded away from the origin whose boundary has measure 0, and $|\cdot|$ is a norm in $\mathbb{R}^d$. An equivalent definition requires an existence of a sequence $b_n \to \infty$ such that

$$nF(b_n\cdot) = n \Pr\left[\frac{\mathbf{X}}{b_n} \in \cdot\right] \xrightarrow{v} \mu(\cdot), \tag{4.2}$$

converges vaguely to μ. We call $\mu(\cdot)$ the *tail measure* of F.

Notice that the special case of $d = 1$ refers to the univariate case, i.e.,

$$\Pr(X > x) \sim x^{-\alpha} L(x), \tag{4.3}$$

where $\alpha > 0$ and L is a slowly varying function (see Section 3.3.1 for the definition) and $\mu(dx) = \alpha x^{-\alpha-1} dx$. The tail index α equals to $1/\xi$, where $\xi > 0$ is the shape parameter used in the previous chapter and can be easily computed by using Hill estimator or MLE where appropriate.

We can consider regular variation in another aspect by transforming (4.2) to the polar coordinate representation. For each $\mathbf{X} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$, let $(r, \mathbf{\Theta}) = (||\mathbf{X}||, \frac{\mathbf{X}}{||\mathbf{X}||})$ and $\mathbb{S} = \{\mathbf{Y} \in \mathbb{R}^d : ||\mathbf{Y}|| = 1\}$, the unit sphere centered at the origin in $\mathbb{R}^d$, then the equation (4.2) is equivalent to

$$n \Pr\left[\left(\frac{r}{b_n}, \mathbf{\Theta}\right) \in \cdot\right] \xrightarrow{v} \mu = c\mu_\alpha \times S, \tag{4.4}$$

for some $c, \alpha > 0$. S is a probability measure on $\mathbb{S}$ and we call S *spectral measure* (or *angular measure* for the case $d = 2$).

We can use the concept from spectral measure (or angular measure) to quantify extreme dependence. We illustrate how we use spectral measure to evaluate the relationship of the extremes by two basic examples: asymptotically independent, and fully dependent.

For asymptotic independence, suppose that $\mathbf{X} = (X^{(1)}, ..., X^{(d)})$ is an iid random vector in $\mathbb{R}^d_+$ with a common distribution F. Therefore, as $x \to \infty$,

$$\bar{F}^{(j)}(x) = \bar{F}(x) = \Pr[X^{(1)} > x] \sim x^{-\alpha} L(x) \tag{4.5}$$

for some $\alpha > 0$ and a slowly varying function L. That is, the vector $\mathbf{X}_i$ has multivariate regularly varying tail probabilities. If we define $b_n := \left(\frac{1}{1-F}\right)^{\leftarrow}(n)$,

then we have

$$n \Pr\left[\frac{\mathbf{X}}{b_n} \in \cdot\right] \xrightarrow{v} \mu(\cdot), \tag{4.6}$$

where $\mu(dx^{(1)}, ..., dx^{(d)}) = \sum_{j=1}^{d} \epsilon_0(dx^{(1)}) \times ... \times \epsilon_0(dx^{(j-1)}) \times \mu_\alpha(dx^{(j)}) \times ... \times \epsilon_0(dx^{(d)})$ and

$$\epsilon_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise.} \end{cases}$$

If we consider the set of points with at least two coordinates being greater than δ in the tails, i.e.,

$$H_{i,j,\delta} := \{\mathbf{x} : x^{(i)} \wedge x^{(j)} > \delta\},$$

then for any $i \neq j$ and $\delta > 0$, we have

$$\mu(H_{i,j,\delta}) = 0.$$

This tells us that the measure is spread along the axes according to one-dimensional measure μ_α and no mass off the axes. This is the same in polar coordinates by considering each axis as a basis vector and the norm that normalizes each basis vector, i.e., for $j = 1, ..., d$, let

$$e_j := (0, ..., 1, ..., 0) \text{ and } ||e_j|| = 1. \tag{4.7}$$

The measure of the set containing all points off the axes is

$$\mu((\bigcup_{j=1}^{d} \{te_j : t > 0\})^c) = 0. \tag{4.8}$$

We know from (4.4) that $\mu \circ T^{-1} = c\mu_\alpha \times S$, where T is the polar transformation. Hence, we can compute S directly from $S(\cdot) = c\mu \circ T^{-1}((1, \infty] \times \cdot)$ for some $c > 0$.

Thus, for any measurable set A, we have

$$\begin{aligned} S(A) &= \mu\left(T^{-1}(\{(r,\mathbf{\Theta}) : r > 1, \mathbf{\Theta} \in A\})\right) \\ &= \mu\left(\{\mathbf{x} : ||\mathbf{x}|| > 1, \frac{\mathbf{x}}{||\mathbf{x}||} \in A\ \}\right) \\ &= \sum_{j=1}^{d} \mu\left(\{\mathbf{x} : ||\mathbf{x}|| > 1, \frac{\mathbf{x}}{||\mathbf{x}||} \in A\ \} \cap \{te_j : t > 0\}\right) \\ &= \sum_{j:e_j \in A} \mu_\alpha(\{te_j : t > 1\}) \end{aligned}$$

That is, S is concentrated on the basis vectors e_j. However, it is noteworthy that the extreme independence is different from independence in a usual context. In fact, it is possible that we have extreme independence in dependent distributions. For example, if we consider a Gaussian random vector (see Figure 4.2a), we can see that the correlation between two components is high. However, if we look at the extremes, the points are near the axes, and this suggests us that the relationship is weak between the components.

If the masses are not concentrated on the axes, then we have extreme dependence between the variables. Now, we take a look at a toy example which illustrates fully dependence of extreme variables. Suppose $\mathbf{X} = (X^{(1)}, ..., X^{(1)})$ has identical coordinates with a distribution F. Assume $X^{(1)}$ has univariate regularly varying tail probabilities, that is, as $x \to \infty$

$$\bar{F}(x) = \Pr[X^{(1)} > x] \sim x^{-\alpha} L(x),$$

for some $\alpha > 0$. Again, if we define $b_n := \left(\frac{1}{1-F}\right)^{\leftarrow}(n)$ then for any positive continuous function f with a compact support in $[\mathbf{0}, \delta\mathbf{1}]^c$, for some $\delta > 0$, we have

$$n\mathbb{E}f(\frac{\mathbf{X}}{b_n}) = \int f(t,...,t) n \Pr\left[\frac{X^{(1)}}{b_n} \in dt\right] = \int f(t\mathbf{1}) n \Pr\left[\frac{X^{(1)}}{b_n} \in dt\right] \tag{4.9}$$

which converges to $\int f(t\mathbf{1})\mu_\alpha(dt)$. Therefore, μ concentrates on $\{t\mathbf{1} : t > 0\}$; equivalently, in polar coordinates, the mass is concentrated on the vector $\frac{\mathbf{1}}{||\mathbf{1}||}$.

Note that the example above might be misleading us to assume that extreme dependence refers to the masses lying on the diagonal line. In fact, we only need that the masses are away from the axes. Let us take a look at another example. Let Y_1, Y_2 be Pareto random variables with shape parameters 1 and 2, respectively. Consider the following set of random variables: with probability 0.5,

$$V_1 = Y_1 + Y_2 \text{ and } V_2 = \frac{1}{2}Y_1 + Y_2;$$

and with probability 0.5,

$$V_1 = \frac{1}{2}Y_1 + Y_2 \text{ and } V_2 = Y_1 + Y_2.$$

Then we generate $n = 10000$ points and convert them into vectors (V_1, V_2). Figure 4.1 illustrates the scatter plot of V_1 and V_2. We see that extreme points are away from the axes and not necessarily align to the diagonal line.

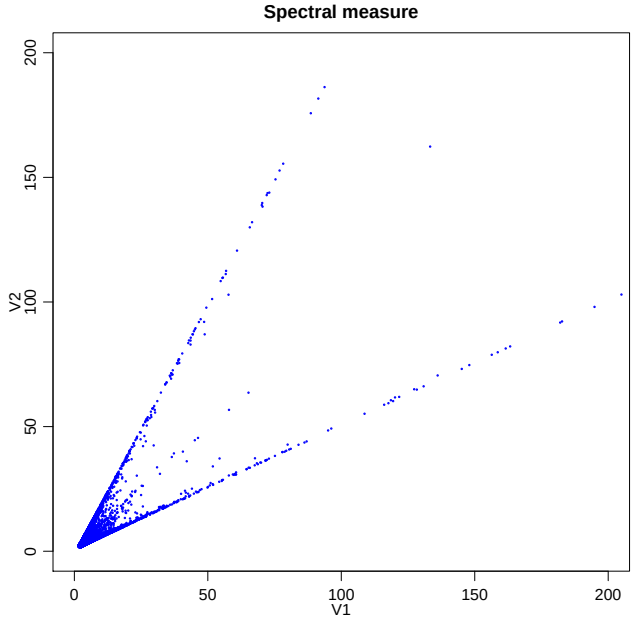


Figure 4.1: An example of extreme dependence. V_1, V_2 are generated by $V_1 = Y_1 + Y_2$ and $V_2 = \frac{1}{2}Y_1 + Y_2$ with probability 0.5, and $V_1 = \frac{1}{2}Y_1 + Y_2$ and $V_2 = Y_1 + Y_2$ with probability 0.5, where $Y_1 = \text{Pareto}(1)$, $Y_2 = \text{Pareto}(2)$ and sample size $n = 10000$.

From the examples above, we can deduce that for 2-dimensional random vectors when the extremes are independent or asymptotically independent, the mass would cluster around the axes. On the other hand, when there are some masses not concentrated around the axes (or 0 and $\pi/2$ if we consider angular measure), this indicates an presence of extreme dependence.

4.3.2 Ranks Method

The tools for measuring the dependence between variables described above only apply to variables with the same marginal tail indices. For real data, this will never be the case. Resnick, (2007)[39] has suggested a number of ways to fix this issue. One is to normalize the marginal indices of each component to 1. Suppose for each $j = 1, ..., d$, the j^{th} component has asymptotically Pareto marginal tail index α_j, then we have

$$\Pr\left[X^{(j)} > x\right] \sim x^{-\alpha_j} \text{ as } x \to \infty. \tag{4.10}$$

Thus, for each j^{th} component, we can take the power of α_j to adjust the tail to have tail index of 1:

$$\Pr[(X^{(j)})^{\alpha_j} > x] = \Pr[X^{(j)} > x^{1/\alpha_j}] \sim x^{-1} \text{ as } x \to \infty. \tag{4.11}$$

With this transformation, we have all random variables with the same tail indices of 1, and hence, we can use spectral measure to analyze the extreme dependence. However, this method generates a lot of errors from each step. Marginal tail indices need to be estimated, and then, we will have to estimate the spectral measure. Each step of estimation can produce errors. At the end, we decide not to go with this method because of possible numerical errors in practice.

Among the methods suggested by Resnick, (2007)[39], we use a transformation that essentially normalizes the tail indices of all components to 1 without calculating or estimating the tail indices α_j for each $j = 1, ..., d$. This method is called *Ranks methods*. The major benefit from this method is that we can avoid the marginal tail index estimation which reduces numerical errors; however, the drawback is that the transformation itself destroys the iid property of the data and makes it more complicated to obtain asymptotic distributions, see Einmahl et al., (2001)[12]. The method can be done as follows.

Let $\mathbf{X}_i = (X_i^{(1)}, ..., X_i^{(d)}), i = 1, ..., n$ be $d-$dimensional vectors. Denote the *rank* of $X_i^{(j)}$ by

$$r_i^{(j)} := \sum_{m=1}^{n} 1_{[X_m^{(j)} \geq X_i^{(j)}]}. \tag{4.12}$$

For a fixed $k > 0$ and for each $i = 1, ..., n$ we transform $\mathbf{X}_i$ into rank vector by

$$\left(X_i^{(1)}, ..., X_i^{(d)}\right) \mapsto \left(\frac{k}{r_i^{(1)}}, ..., \frac{k}{r_i^{(d)}}\right). \tag{4.13}$$

We consider a point $\left(X_i^{(1)}, ..., X_i^{(d)}\right)$ as jointly extreme if $\left|\left|\left(\frac{k}{r_i^{(1)}}, ..., \frac{k}{r_i^{(d)}}\right)\right|\right| > 1$, where $||\cdot||$ is a norm in $\mathbb{R}^d$. In this case, we use the L^2-norm. We use the transformed vectors to estimate the spectral measure.

4.3.3 Estimating Angular Measure

We have been developing the tools to understand tail distributions. In this section, we will provide a way to estimate the spectral measure from the data in polar coordinates.

Since empirical measures (weakly) converge to Poisson random measures, we can use this to estimate spectral measures. Recall the definition of multivariate

regular variation in (4.2); the sequence b_n can be replaced by $b(\frac{n}{k})$, where $k := k(n) \to \infty$ and $n/k \to \infty$. Then, we have the convergence to the tail measure μ, i.e.,

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\mathbf{X}_i/b(\frac{n}{k})} \Rightarrow \mu. \tag{4.14}$$

Similarly, we can use this convergence in polar coordinates as well:

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{(r_i/b(\frac{n}{k}),\mathbf{\Theta}_i)} \Rightarrow c\mu_\alpha \times S. \tag{4.15}$$

In polar coordinates, the factor b in (4.15) is not relevant in the spectral measure S. Thus, we can estimate the spectral measure S by

$$\tilde{S}(\cdot) = \frac{\sum_{\text{extreme points}} \epsilon_{\mathbf{\Theta}}(\cdot)}{\#\text{ of extreme points}}. \tag{4.16}$$

If we use the transformation from the ranks method described in Section 4.3.2, then we consider points outside the unit circle to be extreme. That is, we transform points $\left(\frac{k}{r_i^{(1)}}, ..., \frac{k}{r_i^{(d)}}\right)$ to polar coordinates $(R_i, \mathbf{\Theta}_i)$ and then estimate the empirical spectral measure $\hat{S}$ by using

$$\hat{S}(\cdot) = \frac{\sum_{i=1}^{n} \epsilon_{(R_i,\mathbf{\Theta}_i)}((1,\infty] \times \cdot)}{\sum_{i=1}^{n} \epsilon_{R_i}((1,\infty])}. \tag{4.17}$$

We may notice that the choice of k has a major role on how we categorize extreme points. The higher k is, the more points would lie outside the unit circle, and hence, the more extreme points. We will use an extended version of (3.12) and (3.13) to estimate k; see Nguyen and Samorodnitsky (2013)[35]. The method basically picks the minimum k among the k's from each component. Once we obtain the constant k, we can proceed to estimate the angular measure.

As we pointed out at the end of Section 4.3.1, when extreme dependence of two variables appears, the angular measure tends to have some masses deviated from 0 and $\pi/2$, and when the two variables have extreme independence, the angular

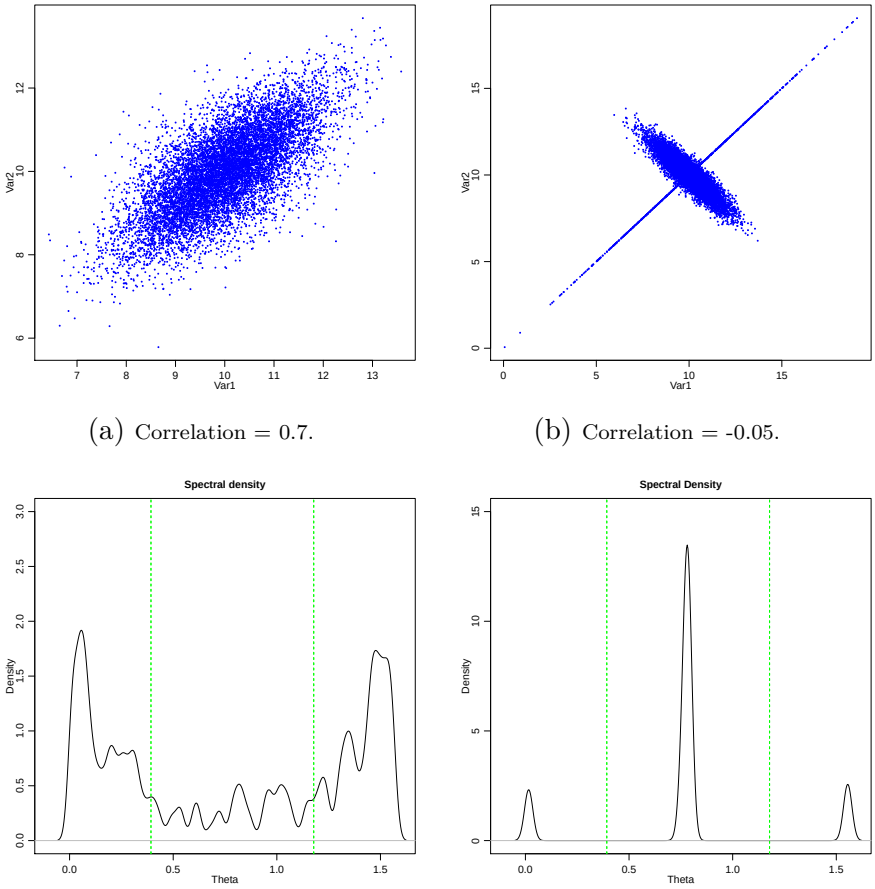
measure would cluster around 0 and $\pi/2$. Thus, an intuitive way to quantify the dependency between two variables is to look at the estimated angular measure and compare the mass of the 'middle' part and the 'outer' parts of the plots. We may want to use kernel density estimation to smooth the empirical angular measure from (4.17).

Since the spectral measure is normalized (i.e., the area under curve from 0 to $\frac{\pi}{2}$ is 1), we can only consider the area of the 'middle' part, which we define to be between $\frac{\pi}{8}$ and $\frac{3\pi}{8}$. For simplicity, we decide to use the angular measure of $[\frac{\pi}{8}, \frac{3\pi}{8}]$, which is approximately equivalent to the area under curve from $\frac{\pi}{8}$ to $\frac{3\pi}{8}$, to represent extreme dependence. Denote this amount by φ:

$$\varphi := \hat{S}([\frac{\pi}{8}, \frac{3\pi}{8}]) \approx \text{area}[\frac{\pi}{8}, \frac{3\pi}{8}], \tag{4.18}$$

where the area is defined in a notion of kernel density estimation from the spectral measure. The range of φ is $[0, 1]$, where $\varphi = 1$ implies extreme dependence and extreme independence implies $\varphi = 0$.

The value of φ is just one way to quantify extreme dependence from the angular measure. In earlier analysis, we also tried to use excess kurtosis of the kernel estimation from angular measure to represent the extreme dependence and the results are consistent with φ. We propose this procedure because it is simpler to program and more intuitive yet with the same accuracy. Note that from the examples in Section 4.3.1, we stated that correlation is not equivalent to extreme dependence without a concrete measurement scale. Now, we have enough tools to quantify the extreme dependence and are able to compare the differences between correlation and extreme dependence. Figure 4.2 shows different scenarios of correlation and extreme dependence which are not necessarily the same. The left panel of Figure 4.2 shows the scenario of the data with high correlation, yet low



(a) Correlation = 0.7.

(b) Correlation = -0.05.

(c) Estimated spectral density. $\varphi = 0.22$.

(d) Estimated spectral density. $\varphi = 0.75$.

Figure 4.2: Examples that shows correlation and estimated spectral measure are not necessarily the same. The plots on the left column (a,c) use the data generated by Gaussian random vectors with correlation $\rho = 0.7$ and each component has $n = 10000$ points sampled from $N(10, 1)$. The data are moderately correlated, while it has low extreme dependence (true $\varphi = 0$; estimated $\varphi = 0.223$). The plots on the right column (b,d) use the data generated by $(Var_1, Var_2) = (Y_1, Y_2)$ with probability 0.8 and $(Var_1, Var_2) = (Z, Z)$ with probability 0.2, where $(Y_1, Y_2) \sim N\left(\boldsymbol{\mu} = \begin{bmatrix}10\\10\end{bmatrix}, \Sigma = \begin{bmatrix}1 & -0.9\\-0.9 & 1\end{bmatrix}\right)$ follows a bivariate normal and $Z \sim N(\mu = 10, \sigma^2 = 9)$. The sample size is also $n = 10000$. The plots show the existence of tail dependence by having masses of extreme away from the axes ($\varphi = 0.75$); however, low correlation ($\rho = -0.05$).

extreme dependence. Figure 4.2a shows the scatter plot of the data and we see correlation in the middle part but when we go to right tail of the distribution, the data do not show such relationship. In contrast, Figure 4.2b shows an example of the data with low correlation but for extreme points, they are highly dependent. The estimated spectral measure of extreme points from these examples are shown in Figures 4.2c-4.2d.

4.4 Discussions and Results

In this discussion, we compare the measurements from CASTNET to the models. We use the same CASTNET data from 25 sites over the continental U.S. as we did in the previous chapter, but for this chapter, we combine the observations into one period and trim the data down to 20 years (1992-2011). To make it easier to visualize the comparisons between CASTNET and each model, Section 4.4.1 - Section 4.4.4 use the considered values from the models as a background and emphasize the values from CASTNET sites by diamonds. The discussion will be as follows. Sections 4.4.1-4.4.2 use the conventional methods, such as means and correlations, to compare CASTNET and the models. Section 4.4.3 uses the methods from the previous chapter, i.e., return levels, to specify the differences in marginal extremes between the observations and the models. Section 4.4.4 will use the methodology we developed in Section 4.3 to quantify extreme dependence between temperature and ozone, and again, we compare these quantities between CASTNET and the models.

4.4.1 Comparison of Mean Quantities

We process the data from the three models by using MDA8 ozone and daily maximum temperature. Figure 4.3 depicts the means of ozone and temperature from each model and those of corresponding locations from CASTNET data represented by diamonds on each plot. The summary of how well the models can simulate data compared to CASTNET is reported in Table 4.1 in the form of correlations of ozone and temperature means between CASTNET sites and model grid points corresponding to those sites.

Table 4.1: Correlations of ozone means and temperature means between CASTNET sites and model simulations. The ozone used in this table is from MDA8 during summertime (1992-2011 for CASTNET data, 1992-2010 for SDM, 2006-2025 for GCM 2000, and 2106-2125 for GCM 2100), and the temperature measurements used in this table are daily maximum during summertime (similar time frame as ozone); neither are deseasonalized.

Model	Ozone	Temperature
Specified Dynamics	0.24	0.57
GCM 2000	0.23	0.53
GCM 2100	0.17	0.53

Since the correlations between the ozone data from simulations and the observed data are considerably low, we would like to look into details of these simulations to find out the differences. Note that the models use temperature as an independent variable which is usually collected from the real observations. Thus, the correlations in temperature are relatively high. Table 4.2 reports biases of ozone and temperature means between models and CASTNET measurements including regional average biases and overall average bias, indicated under each block of CASTNET sites.

Table 4.2: Biases of temperature ($^{\circ}C$) and ozone (ppb) from each model. The biases are based on CASTNET measurements: $Bias = mean(X) - mean(\text{CASTNET})$, where X is either temperature or ozone from the models. We use MDA8 ozone data and daily maximum temperature in the calculation. These measurements do not get deseasonalized, i.e., they still include seasonal cycles. For CASTNET data, we use the measurements from 1992-2013 summertime. For each model, we use summertime data with different years: SDM = 1992-2010, GCM 2000 = 2006-2025, GCM 2100 = 2106-2125.

Site name	State	SDM		GCM 2000		GCM 2100	
		Ozone (ppb)	Temp ($^{\circ}C$)	Ozone (ppb)	Temp ($^{\circ}C$)	Ozone (ppb)	Temp ($^{\circ}C$)
Ashland	ME	12.1	-2.53	14.83	0.01	19.25	2.87
Howland	ME	18.58	-4.77	21.27	0.86	25.62	3.07
Woodstock	NH	19.3	-2.6	25.74	-0.02	30.56	2.78
Connecticut Hill	NY	13.16	-1.27	21.19	1.78	25.98	4.34
Penn State	PA	13.14	-2.78	22.46	0.44	27.11	2.89
Parsons	WV	7.81	-5.12	20.6	-0.8	24.96	1.58
Average	Northeast	14.02	-3.18	21.02	0.38	25.58	2.92
Edgar Evins	TN	13.65	-2.74	24.54	0.12	27.49	2.27
Georgia Station	GA	6.64	-2.78	17.57	-0.25	19.84	1.49
Sand Mountain	AL	7.47	-3.36	18.43	-0.22	21.03	1.72
Candor	NC	11.84	-2.11	21.7	-0.34	24.71	1.54
Coweeta	NC	19.68	-1.62	32.61	2.07	35.72	4.02
Cranberry	NC	6.03	1.63	19.28	5.95	22.88	8.21
Beaufort	NC	15.21	-1.26	23.95	0.65	25.7	2.07
Prince Edward	VA	16.13	1.18	24.7	3.41	28.25	5.49
Shenandoah NP	VA	23.92	-3.41	32.67	-1.4	36.57	0.87
Horton Station	VA	4.79	2.98	16.54	6.14	20.38	8.3
Sumatra	FL	-17.63	6.08	-12.34	6.61	-11.66	7.65
Average	Southeast	9.79	-0.49	19.97	2.07	22.81	3.97

Table 4.2: (Continued).

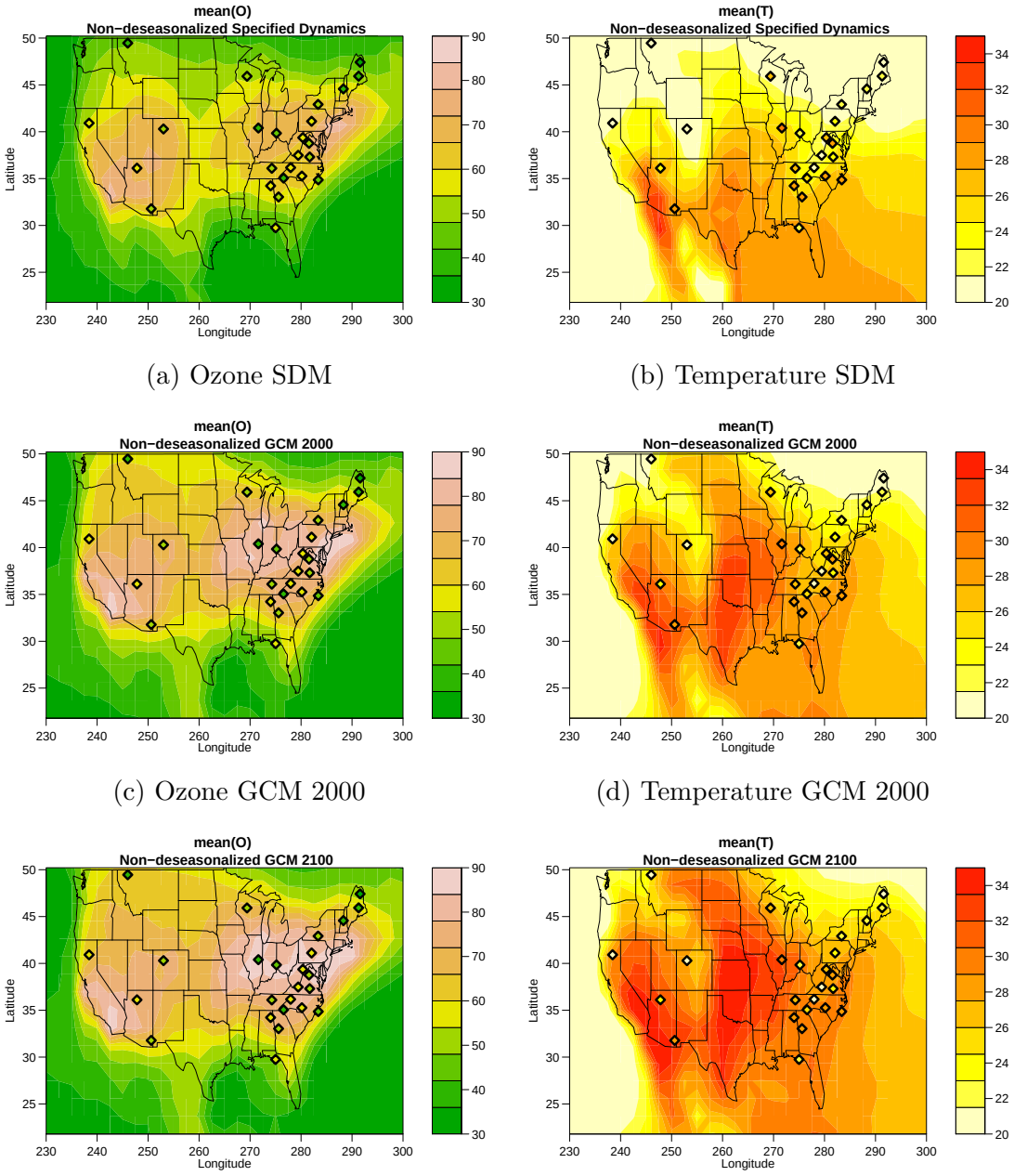
Site name	State	SDM Ozone (ppb)	SDM Temp ($^{\circ}C$)	GCM 2000 Ozone (ppb)	GCM 2000 Temp ($^{\circ}C$)	GCM 2100 Ozone (ppb)	GCM 2100 Temp ($^{\circ}C$)
Bondville	IL	29.46	-4.93	42.54	-0.82	46.27	1.66
Perkinstown	WI	2.47	-4.35	14.99	-0.24	18.96	2.35
Oxford	OH	29.3	2.66	40.76	6.19	43.46	8.51
Average	Midwest	20.41	-2.21	32.76	1.71	36.23	4.17
Glacier NP	MT	11.4	-6.1	23.48	-0.7	26.1	2.6
Chiricahua	AZ	5.8	-0.36	9.89	3.58	8.65	5.32
Grand Canyon	AZ	14.78	1.32	18.55	5.12	18.98	7.4
Gothic	CO	14.6	3.62	17.63	9.25	18.31	11.46
Lassen Volcanic	CA	-0.7	-0.44	4.79	4.49	6.75	7.43
Average	West	9.18	-0.39	14.87	4.35	15.76	6.84
Average	All	11.96	-1.32	20.73	2.08	23.67	4.32

From Table 4.2, western U.S. has the least ozone bias but the highest temperature bias among all regions for all three models. Midwest has the highest ozone bias and moderate temperature bias. Eastern U.S have relatively low biases for all three models. In general, SDM has the least bias among the three models and GCM 2100 has the largest bias among these models.

4.4.2 Comparison of Conditional Quantities

Now, we continue examining the means, and correlations of ozone and temperature for each model. In many cases, we may prefer to consider properties conditional on high temperature and/or ozone rather than unconditional ones. We focus on the analysis conditionally on high temperature because in general the ozone level is affected by temperature. We have seen the unconditional ozone and temperature means in Figure 4.3 and in contrast, Figure 4.4 shows ozone means conditional on different temperature quantiles from Specified Dynamics Model (GCM 2000 and GCM 2100 give similar results and we opt out the results here). Recall the normalized scale data from (2.2) and rescaled data from (3.4) in previous chapters that we transform the data to get rid of seasonal cycles and year-to-year effects (normalized scale) and then reverse back to original scale by using average of means and standard deviations (rescaled). Figure 4.4 uses the rescaled ozone (ppb) instead of normalized scale (no unit) and the pattern of the ozone average does not change much by using different temperature thresholds. This indicates that the ozone level only slightly increases as temperature gets higher.

In contrast to Figure 4.4, Figure 4.5 shows the ozone means conditional on different thresholds of temperature in the normalized scale. The figure shows the changes in the relationship between extreme temperature and ozone as the



(a) Ozone SDM

(b) Temperature SDM

(c) Ozone GCM 2000

(d) Temperature GCM 2000

(e) Ozone GCM 2100

(f) Temperature GCM 2100

Figure 4.3: Left column (a),(c),(e): Ozone averages (ppb) from SDM, GCM 2000 and GCM 2100, respectively. Right column (b),(d),(f): Temperature averages ($^{\circ}C$) from SDM, GCM 2000 and GCM 2100, respectively. The averages are calculated from MDA8 (for ozone) and daily maximum (for temperature) without deseasonalization. The diamonds are averages of ozone (or temperature) from CASTNET sites.

threshold increases. We can also notice major differences between Eastern U.S. and Western U.S. As we increase the temperature thresholds to 90^{th} percentile or higher, the conditional ozone means in normalized scale are still low in Western U.S. (around 0.1-0.5). This implies low extreme dependence between temperature and ozone. However, the results from Eastern U.S. contrast the relationship from Western U.S. in that the conditional ozone level in normalized scale is much higher (about 1.2-1.5).

To emphasize on the different scale usage, the normalized scale better represents conditional ozone means in the sense that it shows how high the ozone is, relative to its own location. For example, if we have an ozone level of 80 ppb in original scale at location X, then at this level, it may equal to 2 in normalized scale. While at location Y, the same ozone level of 80 ppb could be equivalent to only 1 in normalized scale because other data points at this location also have high ozone levels. Thus, the normalized scale carries more information over the original scale. Combining the results from Figure 4.4 and Figure 4.5, they show some dependencies between extreme temperature and extreme ozone geographically for each model.

Next, we will consider the conditional correlation between temperature and ozone from each model. We note that this is not the correlation between models. Figure 4.6 shows the unconditional (Figure 4.6a) and conditional (Figures 4.6b-4.6f) correlations between ozone and temperature at different thresholds of temperature based on quantiles. From the unconditional correlation map (Figure 4.6a), we see that ozone and temperature are highly correlated especially for continental U.S. However, the bigger temperature threshold is, the less consistency in conditional correlation between ozone and temperature becomes. This behavior

can be observed from 4.6b-4.6f. In particular, the Figure 4.6f shows very scattered and random correlations between ozone and temperature, such behavior even happens in the oceanic areas. This suggests either that the number of data is too small or that once the temperature moves into higher zone, the linear relationship between temperature and ozone declines. Figure 4.7 shows the p-values of the changes in correlations from one threshold to another, corresponding to Figure 4.6. Only moving from unconditional to conditional on temperature being greater than 80% shows the significant change here. Even the changes in correlations from 80% to 95% are still not significant (p-values are greater than 0.1). This shows that it is not practical to pick the exact threshold to consider particular values as extreme.

Tables 4.3-4.5 show the differences in conditional ozone means between the models and CASTNET. That is, for Table 4.3 and Table 4.4 and for each location of CASTNET sites, the difference is defined by

$$\begin{aligned}\Delta\text{mean} = &\text{ mean(Model } O_3|\text{Model } T > x\%) \\ &- \text{mean(CASTNET } O_3|\text{CASTNET } T > x\%),\end{aligned}$$

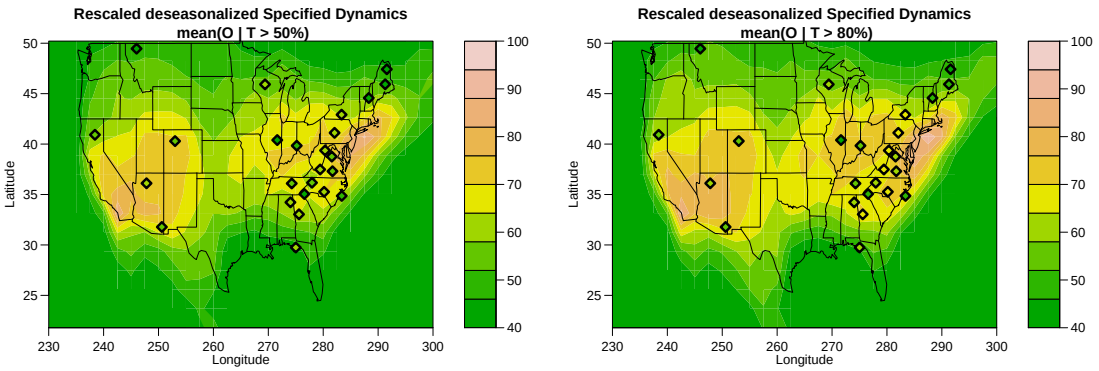
and Table 4.5 shows the differences in conditional correlations:

$$\begin{aligned}\Delta\text{correlation} = &\text{ cor(Model } T, O_3|\text{Model } T > x\%) \\ &- \text{cor(CASTNET } T, O_3|\text{CASTNET } T > x\%)\end{aligned}$$

where $x = 80, 90, 95$. We consider these differences as biases of the models. We group these CASTNET sites by regions and report regional and overall differences as well. For the differences in conditional means using rescaled data (Table 4.3), all models have positive biases compared to CASTNET data. In general, GCM 2100 has the most biases among the models and SDM has the least biases. However, if we

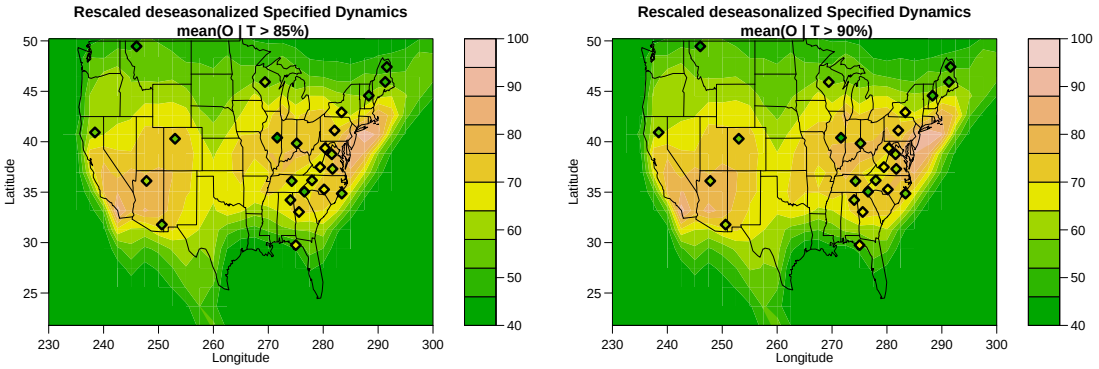
consider the differences in conditional means using the normalized scale (Table 4.4), then most sites in northeastern U.S. have negative biases compared to CASTNET data but the other regions of the U.S. still have positive biases similar to rescaled data. For the differences in conditional correlations using the normalized scale (Table 4.5), only southeastern U.S. has relatively higher biases compared to other regions for all models.

Since there is a large discrepancy between conditional correlations from the models and that from CASTNET measurements, we would like to see how well the models can simulate extreme climates. Let us take a look at the differences across the models. Figure 4.8a shows a comparison between the mean of extreme temperature at each location from specified dynamics model minus means of extreme temperature from GCM 2000 in the normalized scale: mean(SDM T|SDM $T > 90\%$)-mean(GCM2000 T|GCM2000 $T > 90\%$). Since the specified dynamics model covers the data from 1992-2010 and GCM 2000 covers the data from 2006-2025, it is unfair to compare the data with different time frame. Thus, we only take the overlapped years between these two models (2006-2010). Again, we define 'extreme points' to be points corresponding to temperature at 90^{th} quantile or higher. Similarly, we do the same procedure to GCM 2100 and GCM 2000. Figure 4.8c shows the mean of extreme temperature from GCM 2100 minus the mean of extreme temperature from GCM 2000 in the normalized scale. Figure 4.8b-4.8d are performed in similar ways except that we use extreme ozone instead. We also checked the p-values corresponding to Figure 4.8 on the null hypothesis of the differences are zero but the p-values are greater than 0.1 for all grid points. That is, we do not have enough evidence to say conclude that the differences in the normalized scale among the models are significantly non-zero.



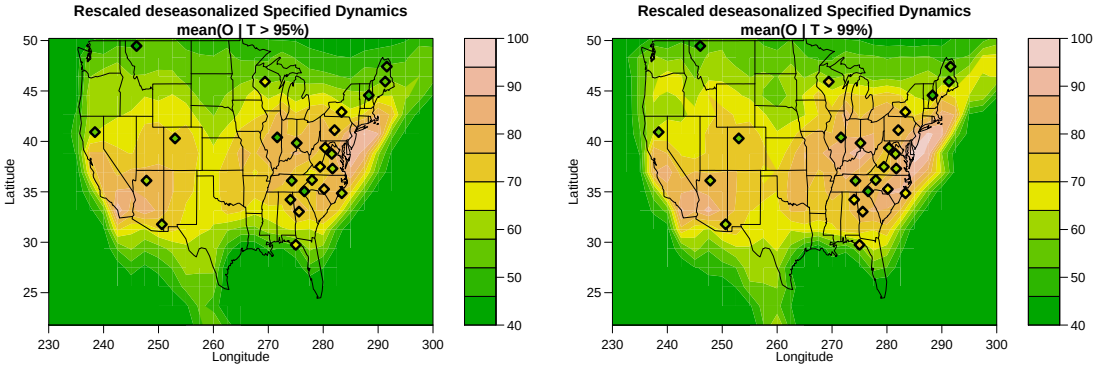
(a) Threshold = 50%

(b) Threshold = 80%



(c) Threshold = 85%

(d) Threshold = 90%



(e) Threshold = 95%

(f) Threshold = 99%

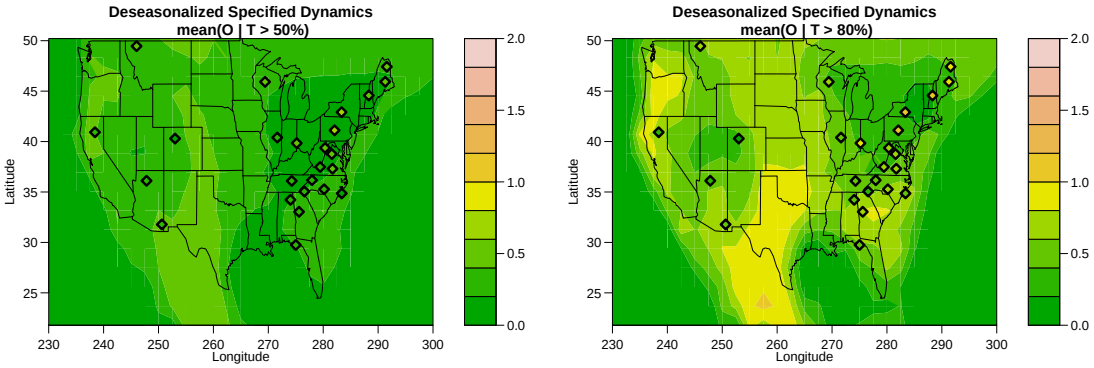
Figure 4.4: Averages of rescaled ozone (ppb) conditional on certain levels of temperature from Specified Dynamics model. The thresholds are used in percentile of temperature indicated by the numbers under each plot. Other models have similar results. We use MDA8 ozone and daily maximum temperature summertime (1992-2010) in the calculation.

Table 4.3: Biases (ppb) of ozone means conditional on different levels of temperature thresholds on each model. The biases are based on CASTNET measurements: $Bias = mean(\text{Model } O_3|\text{Model } T > x\%) - mean(\text{CASTNET } O_3|\text{CASTNET } T > x\%)$, where $x = 80, 90, 95$. We use MDA8 ozone data and daily maximum temperature in the calculation. These measurements are deseasonalized and rescaled back to original scale, see (3.4) for rescaling procedure. For CASTNET data, we use the measurements from 1992-2013 summertime. For each model, we use summertime data with different years: SDM = 1992-2010, GCM 2000 = 2006-2025, GCM 2100 = 2106-2125.

Site name	State	SDM			GCM 2000			GCM 2100		
		80%	95%	99%	80%	95%	99%	80%	95%	99%
Ashland	ME	8.21	4.18	6.07	10.26	4.23	6.74	15.06	8.53	10.48
Howland	ME	14.6	10.19	19.55	16.02	10.48	17.19	20.68	14.59	18.6
Woodstock	NH	10.76	7.42	11.52	15.05	12.49	13.99	20.71	15.69	16.65
Connecticut Hill	NY	3.17	3.55	12.02	9.43	7.37	12.5	14.86	12.61	11.57
Penn State	PA	1.17	-1.06	-0.42	7.63	1.4	2.96	14.4	9.64	8.55
Parsons	WV	0.54	2.8	8.3	12.88	13.79	19.41	20.15	20.71	26.44
Average	Northeast	6.41	4.51	9.51	11.88	8.29	12.13	17.64	13.63	15.38
Edgar Evins	TN	11.62	17.06	21.39	24.03	28.79	29.74	29.04	31.67	34.81
Georgia Station	GA	6.09	9.57	5.08	15.73	15.9	18.8	19.68	21.44	18.86
Sand Mountain	AL	7.03	12.19	12.79	19.11	24.07	25.68	23.48	26.59	26.01
Candor	NC	10.37	9.55	22.35	19.75	16.95	21.95	24.18	21.09	24.44
Coweeta	NC	21.38	23.81	27.84	35.84	35.73	40.29	40.6	40.3	42.66
Cranberry	NC	3.33	6.03	13.95	19.01	21.39	26.07	25.02	25.45	30.98
Beaufort	NC	19.3	17.08	23.23	29	23.22	23.56	31.68	24.77	25.02
Prince Edward	VA	14.89	18.76	22.01	23.72	25.7	25.13	28.79	31.3	30.87
Shenandoah NP	VA	19.88	20.29	24.31	26.19	23.69	27.12	32.5	31.85	35.23
Horton Station	VA	2.38	3.65	10.55	15.15	15.45	22.04	21.21	21.63	27.73
Sumatra	FL	-18.57	-17.83	-14.7	-12.43	-12.82	-11.1	-12.05	-14.09	-12.82
Average	Southeast	8.88	10.92	15.35	19.55	19.82	22.66	24.01	23.82	25.8

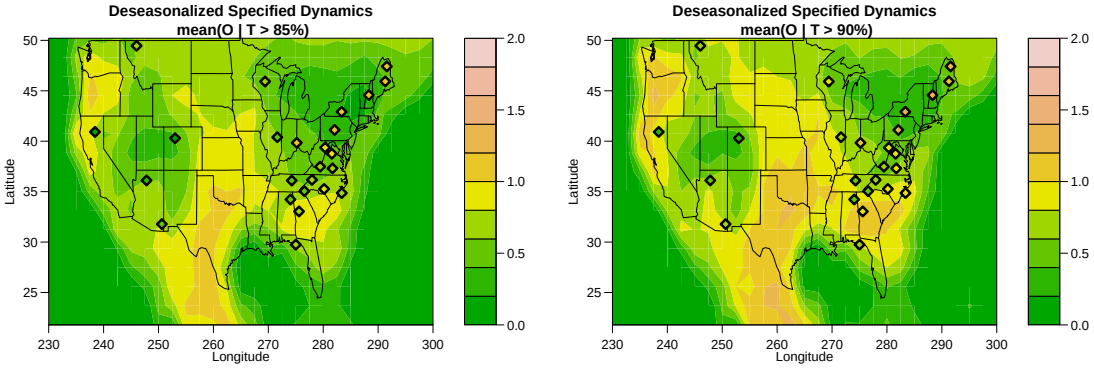
Table 4.3: (Continued).

Site name	State	SDM			GCM 2000			GCM 2100		
		80%	95%	99%	80%	95%	99%	80%	95%	99%
Bondville	IL	28.16	29.46	26.83	40.39	38.71	31.15	45.29	43.02	37.08
Perkinstown	WI	-1.46	-2.6	-3.85	10.15	8.78	3.22	12.69	10.75	2.73
Oxford	OH	23.27	25.96	31.76	33.81	32.85	29.99	37.76	34.44	34.74
Average	Midwest	16.66	17.61	18.25	28.12	26.78	21.45	31.91	29.4	24.85
Glacier NP	MT	9.99	10.04	9.46	20.05	19.72	21.26	22.18	21.1	17.93
Chiricahua	AZ	10.21	11.44	12.91	12.71	13.57	14.66	11.78	12.82	15.15
Grand Canyon	AZ	16.39	17.99	12.15	18.38	17.62	13.7	18.94	19.18	15.82
Gothic	CO	15.98	15.68	17.68	19.3	20.01	22.14	19.66	20.58	21.35
Lassen Volcanic	CA	4.15	5.99	7.45	6.26	7.98	9.04	8.74	8.94	7.47
Average	West	11.34	12.23	11.93	15.34	15.78	16.16	16.26	16.52	15.54
Average	All	9.71	10.45	13.61	17.9	17.08	18.69	21.88	20.58	21.13



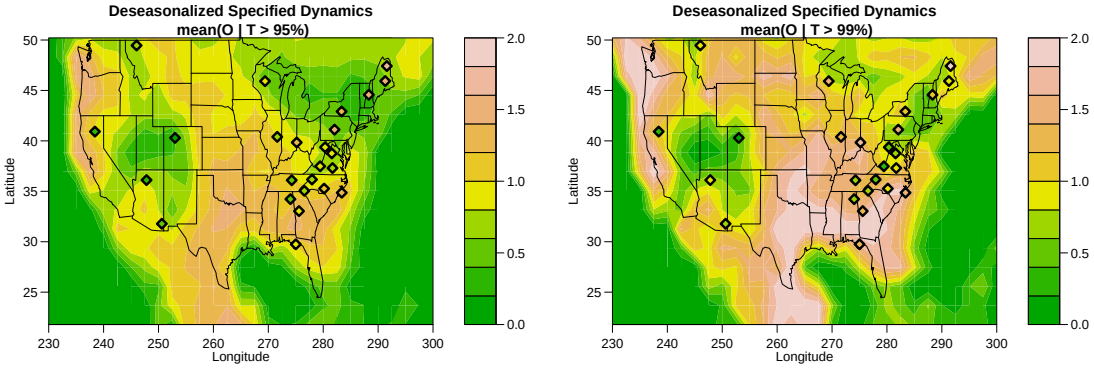
(a) Threshold = 50%

(b) Threshold = 80%



(c) Threshold = 85%

(d) Threshold = 90%



(e) Threshold = 95%

(f) Threshold = 99%

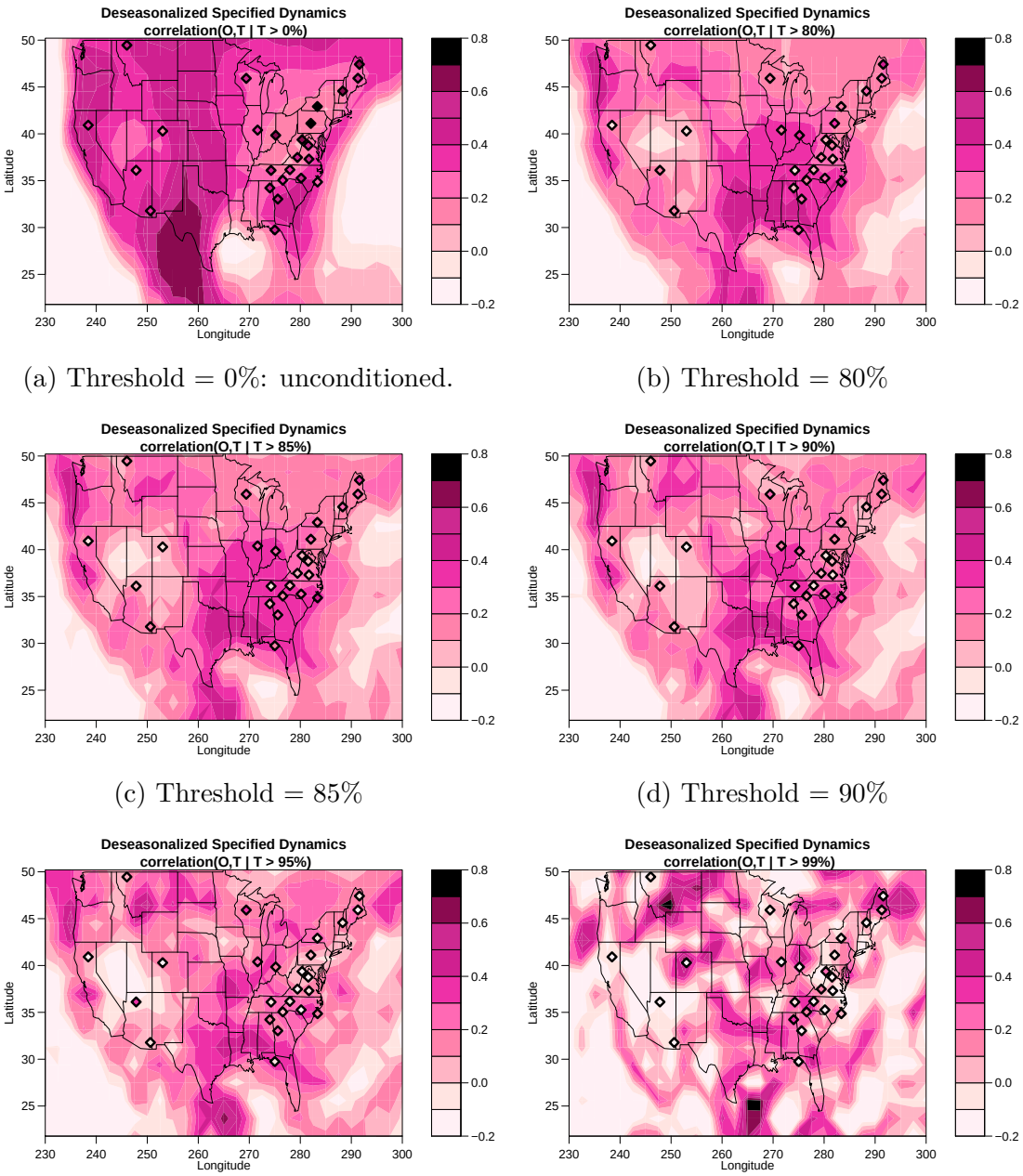
Figure 4.5: Averages of deseasonalized ozone conditioned on high temperature from Specified Dynamics model (normalized scale, and hence no unit). Different levels of percentile are indicated by the number of each plot. Note that these plots are based on the standardized, so the unconditional mean is 0 so we omitted the unconditional means here. The averages of temperature conditioned on high ozone have similar results. Other models (GCM 2000 and GCM 2100) also have similar results.

Table 4.4: Biases of ozone means conditional on different levels of temperature thresholds on each model. Similar to Table 4.3 except that we used deseasonalized ozone and deseasonalized temperature data.

Site name	State	SDM			GCM 2000			GCM 2100		
		80%	95%	99%	80%	95%	99%	80%	95%	99%
Ashland	ME	-0.53	-1.06	-0.91	-0.57	-1.26	-1.05	-0.54	-1.27	-1.11
Howland	ME	-0.45	-0.86	-0.12	-0.5	-0.98	-0.42	-0.5	-1.02	-0.66
Woodstock	NH	-0.87	-1.2	-0.83	-1.04	-1.31	-1.15	-0.97	-1.45	-1.33
Connecticut Hill	NY	-0.83	-0.8	-0.11	-0.97	-1.13	-0.7	-0.92	-1.11	-1.19
Penn State	PA	-0.81	-0.91	-0.86	-1.07	-1.53	-1.37	-0.85	-1.17	-1.27
Parsons	WV	-0.44	-0.15	0.28	-0.48	-0.31	0.14	-0.2	-0.09	0.36
Average	Northeast	-0.66	-0.83	-0.42	-0.77	-1.09	-0.76	-0.66	-1.02	-0.87
Edgar Evins	TN	-0.11	0.44	0.9	0.09	0.6	0.74	0.24	0.5	0.83
Georgia Station	GA	0.12	0.44	0.19	0.23	0.42	0.89	0.32	0.61	0.54
Sand Mountain	AL	0.06	0.54	0.62	0.28	0.83	1.06	0.4	0.72	0.7
Candor	NC	-0.05	-0.08	0.9	0.1	-0.03	0.38	0.17	-0.01	0.24
Coweeta	NC	0.16	0.39	0.76	0.42	0.46	0.95	0.53	0.52	0.75
Cranberry	NC	-0.24	0.06	0.9	0.05	0.34	0.84	0.25	0.29	0.86
Beaufort	NC	0.21	-0.04	0.33	0.43	0	0.04	0.45	-0.08	-0.06
Prince Edward	VA	-0.2	0.09	0.32	-0.06	0.13	0.08	0.03	0.24	0.2
Shenandoah NP	VA	-0.34	-0.32	0.01	-0.48	-0.69	-0.38	-0.26	-0.29	0.02
Horton Station	VA	-0.2	-0.05	0.61	-0.04	0.04	0.7	0.14	0.2	0.78
Sumatra	FL	-0.08	0	0.32	0.22	0.32	0.66	0.26	0.17	0.5
Average	Southeast	-0.06	0.13	0.53	0.11	0.22	0.54	0.23	0.26	0.49

Table 4.4: (Continued).

Site name	State	SDM			GCM 2000			GCM 2100		
		80%	95%	99%	80%	95%	99%	80%	95%	99%
Bondville	IL	-0.05	0.12	-0.03	-0.09	-0.2	-0.81	0.02	-0.14	-0.59
Perkinstown	WI	-0.26	-0.32	-0.34	-0.4	-0.51	-0.98	-0.54	-0.7	-1.37
Oxford	OH	-0.6	-0.39	0.08	-0.6	-0.65	-0.9	-0.48	-0.79	-0.72
Average	Midwest	-0.3	-0.2	-0.1	-0.36	-0.45	-0.9	-0.33	-0.54	-0.89
Glacier NP	MT	-0.12	-0.11	-0.17	-0.34	-0.35	-0.01	-0.44	-0.58	-1.04
Chiricahua	AZ	0.26	0.34	0.45	0.27	0.36	0.47	0.26	0.35	0.59
Grand Canyon	AZ	0.13	0.32	-0.49	0.03	-0.09	-0.66	0.08	0.13	-0.33
Gothic	CO	0.17	0.13	0.44	0.31	0.44	0.79	0.29	0.46	0.55
Lassen Volcanic	CA	0.54	0.74	0.9	0.37	0.68	0.85	0.41	0.47	0.24
Average	West	0.2	0.28	0.23	0.13	0.21	0.29	0.12	0.17	0
Average	All	-0.18	-0.11	0.17	-0.15	-0.18	0.01	-0.07	-0.16	-0.1



(a) Threshold = 0%: unconditioned.

(b) Threshold = 80%

(c) Threshold = 85%

(d) Threshold = 90%

(e) Threshold = 95%

(f) Threshold = 99%

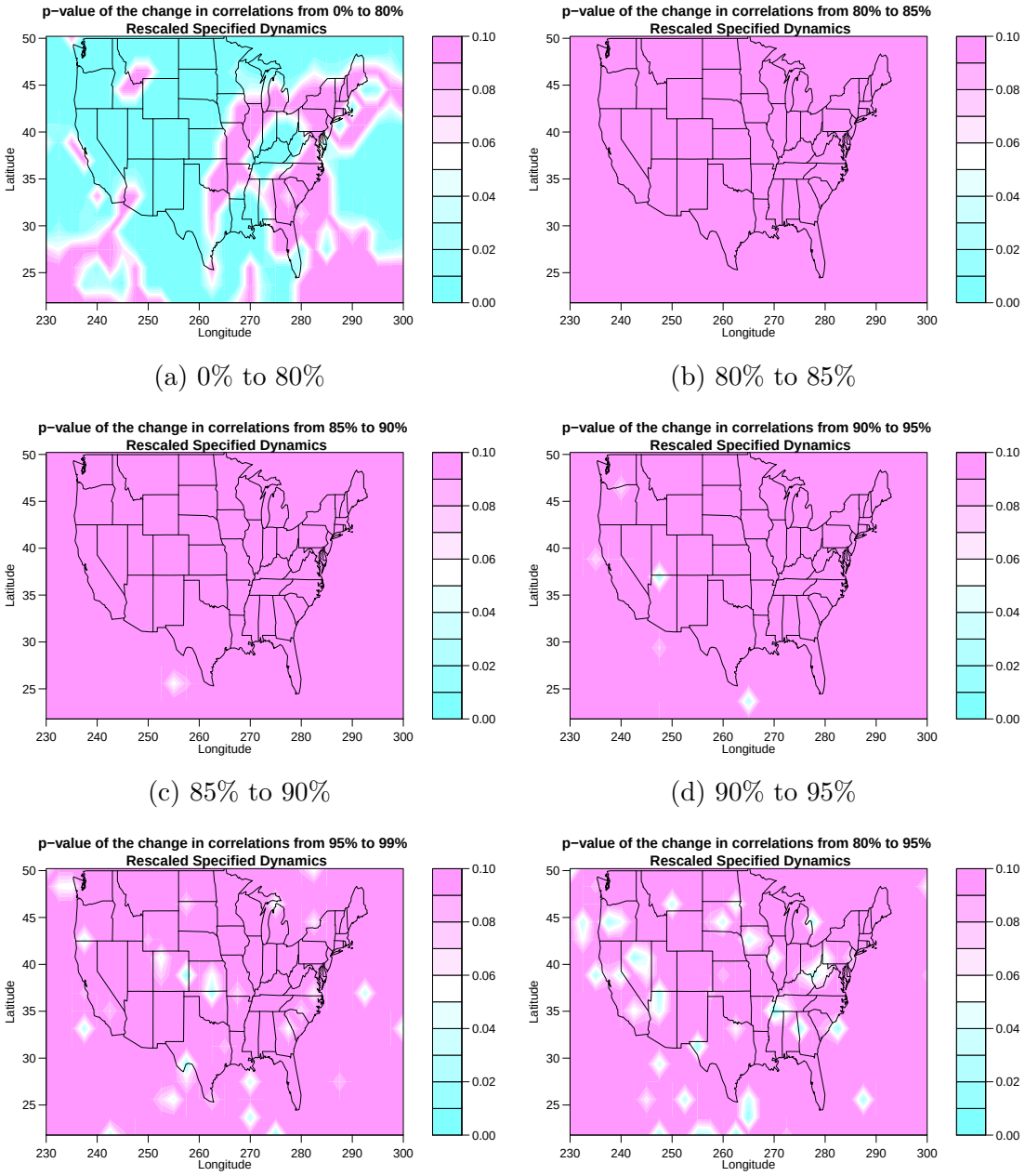
Figure 4.6: Correlations between deseasonalized temperature and deseasonalized ozone conditioned on high temperature from Specified Dynamics model. Different levels of percentile thresholds are indicated by the number of each plot. Other models have similar results. We use MDA8 ozone and daily maximum temperature summertime from 1992-2010 in deseasonalizing the data.

Table 4.5: Biases of correlations between normalized ozone and normalized temperature conditional on different levels of temperature thresholds on each model. The biases are based on CASTNET measurements: $Bias = \text{cor}(\text{Model } T, O_3 | \text{Model } T > x\%) - \text{cor}(\text{CASTNET } T, O_3 | \text{CASTNET } T > x\%)$, where $x = 80, 95, 99$. See (3.3) in Section 3.2 for the deseasonalization procedure. Again, the ranges of the data from each model are different. For CASTNET data, we use the measurements from summertime 1992-2013. For each model, the data are summertime from different years: SDM = 1992-2010, GCM 2000 = 2006-2025, GCM 2100 = 2106-2125.

Site name	State	SDM			GCM 2000			GCM 2100		
		80%	95%	99%	80%	95%	99%	80%	95%	99%
Ashland	ME	-0.2	0.23	0.68	-0.32	0.3	0.8	-0.38	0.19	0.3
Howland	ME	-0.18	0.2	0.06	-0.25	0.15	-0.05	-0.31	0.07	-0.06
Woodstock	NH	-0.2	0.14	0.54	-0.19	-0.1	0.47	-0.3	-0.05	0.48
Connecticut Hill	NY	0.01	0.31	0.1	-0.15	0.14	0.05	-0.18	-0.04	-0.05
Penn State	PA	-0.12	-0.06	0.09	-0.28	-0.01	-0.13	-0.24	-0.15	-0.15
Parsons	WV	0.19	0.21	-0.44	0.13	0.34	-0.02	0.08	0.12	-0.81
Average	Northeast	-0.08	0.17	0.17	-0.18	0.14	0.19	-0.22	0.02	-0.05
Edgar Evins	TN	0.36	0.33	0.39	0.39	0.31	0.25	0.25	0.25	0.14
Georgia Station	GA	0.2	-0.13	0.43	0.17	0.21	0.63	0.29	0.1	0.65
Sand Mountain	AL	0.23	0.07	-0.16	0.32	0.15	-0.25	0.26	0.09	0.06
Candor	NC	0.1	0.5	0.18	0.05	0.34	0.41	-0.03	0.17	0
Coweeta	NC	0.18	0.16	0.38	0.17	0.29	-0.03	0.11	0.13	0.22
Cranberry	NC	0.22	0.38	0.6	0.24	0.26	0.27	0.1	0.26	0.58
Beaufort	NC	-0.01	0.26	0.32	-0.07	0.2	0.61	-0.15	0.01	-0.35
Prince Edward	VA	0.19	0.2	-0.06	0.16	0.14	0.61	0.18	0.15	0.36
Shenandoah NP	VA	0.01	0.21	0.31	-0.15	0.25	0.2	0.01	0.31	0.41
Horton Station	VA	0.1	0.31	-0.13	0.14	0.35	-0.12	0.12	0.22	-0.09
Sumatra	FL	0.12	0.41	0.76	0.2	0.48	1.32	0.06	0.43	0.85
Average	Southeast	0.15	0.25	0.27	0.15	0.27	0.35	0.11	0.19	0.26

Table 4.5: (Continued).

Site name	State	SDM			GCM 2000			GCM 2100		
		80%	95%	99%	80%	95%	99%	80%	95%	99%
Bondville	IL	0.12	0.04	0.32	-0.14	-0.23	-0.13	-0.1	-0.18	-0.11
Perkinstown	WI	-0.02	-0.08	-0.15	-0.11	-0.28	0.41	-0.13	-0.49	-0.47
Oxford	OH	0.08	0.06	0.57	-0.13	-0.1	0.23	-0.2	0.04	0.01
Average	Midwest	0.06	0.01	0.25	-0.13	-0.2	0.17	-0.14	-0.21	-0.19
Glacier NP	MT	0.03	-0.03	-0.06	0.05	0.2	0.3	-0.11	-0.1	0.1
Chiricahua	AZ	0.03	0.1	0.08	0.05	0.19	0.41	0.09	0.23	0.26
Grand Canyon	AZ	-0.04	-0.54	-0.15	-0.19	-0.39	-0.19	-0.06	-0.21	0.43
Gothic	CO	0.09	0.47	1.26	0.13	0.25	0.79	0.12	0.28	0.73
Lassen Volcanic	CA	0.25	0.28	0.38	0.33	0.21	0.28	0.13	0.03	0.41
Average	West	0.07	0.06	0.3	0.07	0.09	0.32	0.03	0.05	0.39
Average	All	0.07	0.16	0.25	0.02	0.15	0.28	-0.02	0.07	0.16



(a) 0% to 80%

(b) 80% to 85%

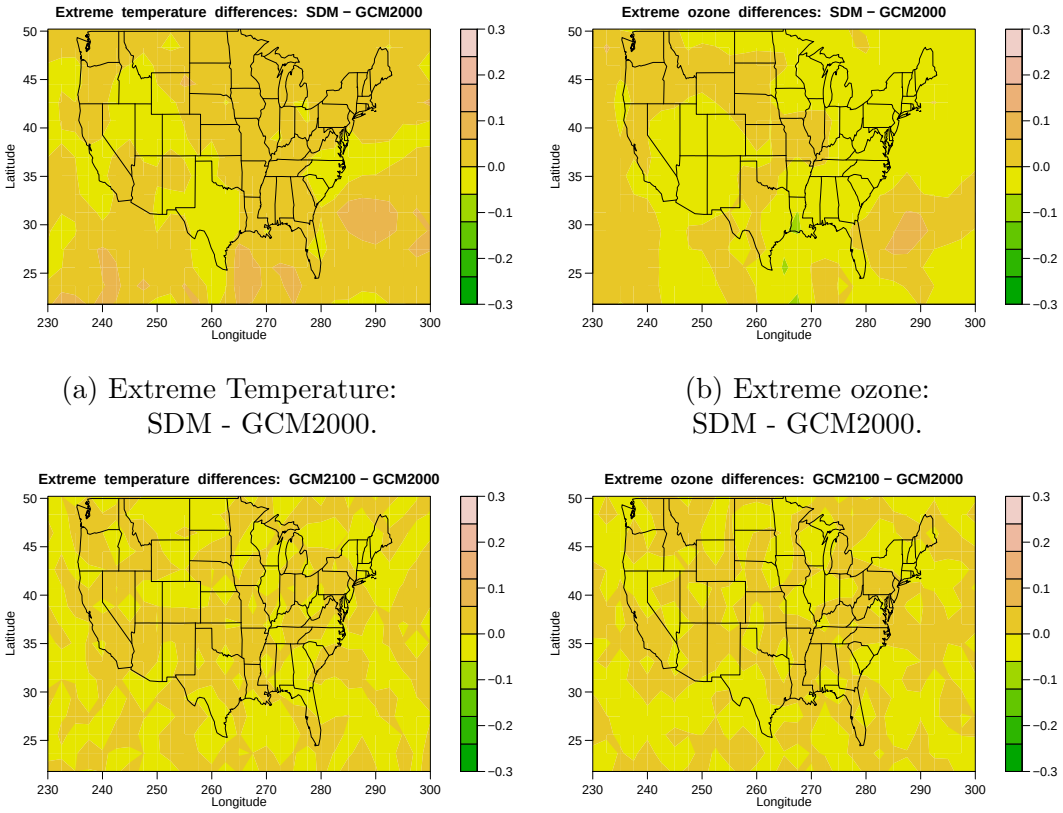
(c) 85% to 90%

(d) 90% to 95%

(e) 95% to 99%

(f) 80% to 95%

Figure 4.7: p-values of the changes in conditional correlations using one threshold to another threshold indicated by the numbers in the caption under each panel. See Figure 4.6 for conditional correlations using different temperature thresholds.



(a) Extreme Temperature: SDM - GCM2000.

(b) Extreme ozone: SDM - GCM2000.

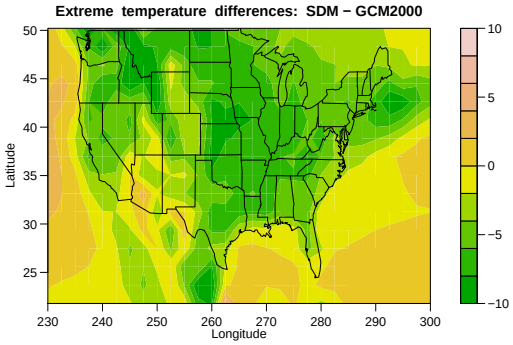
(c) Extreme temperature: GCM2100-GCM2000

(d) Extreme ozone: GCM2100-GCM2000.

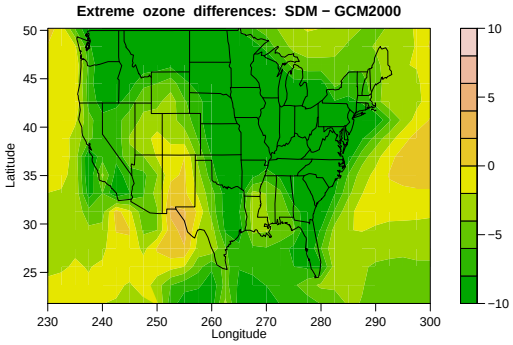
Figure 4.8: Differences across the models of extreme temperature averages and extreme ozone averages in the normalized scale. (a) mean($T_1|T_1 > 90\%$) - mean($T_2|T_2 > 90\%$); (b) mean($O_1|T_1 > 90\%$) - mean($O_2|T_2 > 90\%$); (c) mean($T_3|T_3 > 90\%$) - mean($T_2|T_2 > 90\%$); (d) mean($O_3|T_3 > 90\%$) - mean($O_2|T_2 > 90\%$). The subscripts 1,2, and 3 refer to specified dynamics model, GCM 2000, and GCM 2100, respectively.

Interestingly, if we rescale the data to its original scale ($^{\circ}C$ for temperature and ppb for ozone, see (3.4) for the rescaling procedure), we have different outcomes. Figure 4.9a shows that the rescaled extreme ozone levels from specified dynamics model are usually lower than the extreme ozone from GCM 2000 across continental areas. If we look at all plots in Figure 4.9, we see that the extreme levels of both rescaled temperature and rescaled ozone from specified dynamics are lower than

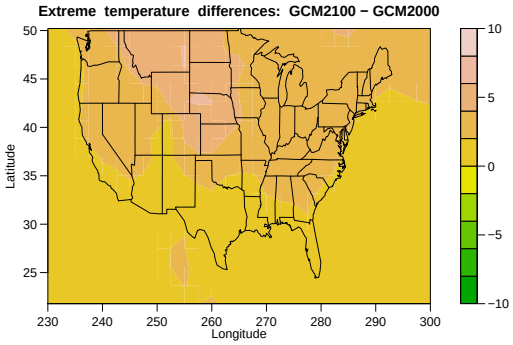
those from GCM 2000. Similarly, those from GCM 2000 are also lower than GCM 2100. We can also confirm the results from Figure 4.10 which shows the p-values on corresponding panels from Figure 4.9 under the null hypothesis of the differences of extreme means from particular model pairs are zero.



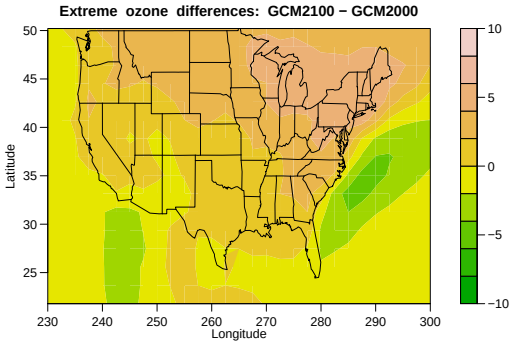
(a) Extreme Temperature: SDM - GCM2000.



(b) Extreme ozone: SDM - GCM2000.

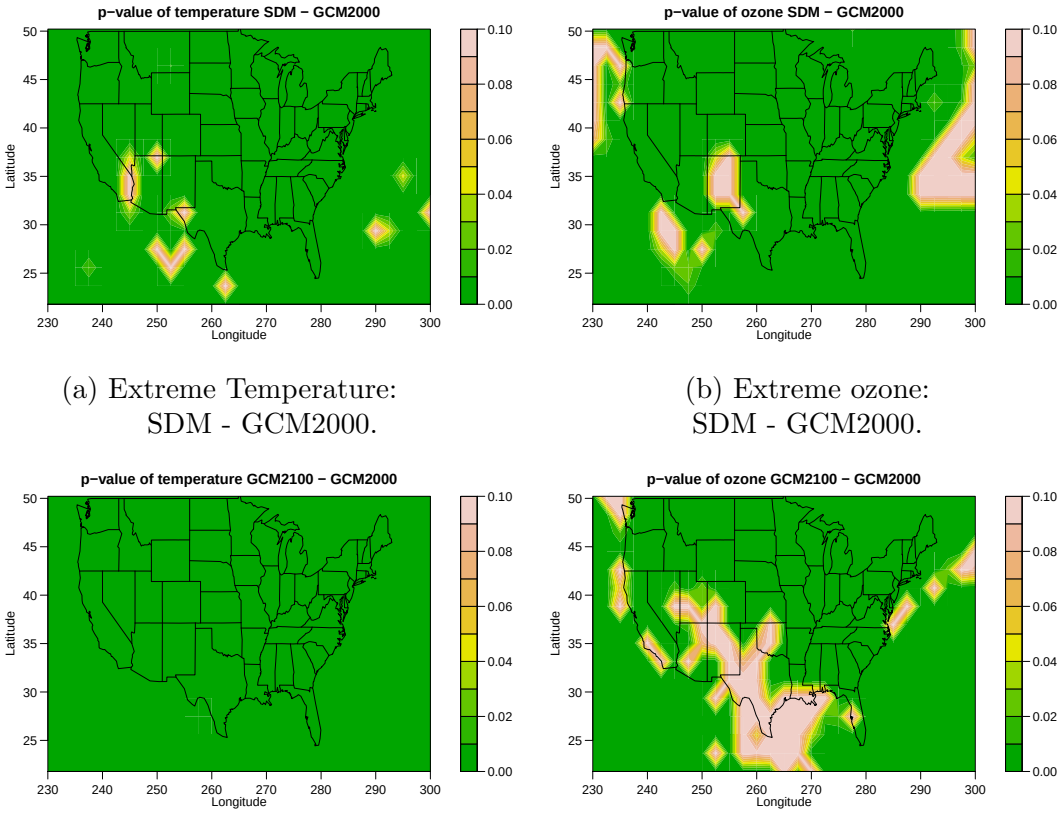


(c) Extreme temperature: GCM2100-GCM2000



(d) Extreme ozone: GCM2100-GCM2000.

Figure 4.9: Similar to Figure 4.8, but using rescaled data (see (3.4) for the rescaling procedure).



(a) Extreme Temperature: SDM - GCM2000.

(b) Extreme ozone: SDM - GCM2000.

(c) Extreme temperature: GCM2100-GCM2000

(d) Extreme ozone: GCM2100-GCM2000.

Figure 4.10: p-values of corresponding panels in Figure 4.9. We are testing on the null hypothesis of the difference in means between particular model pairs is zero (normalized scale). Most grid points have p-values lower than 0.05 which suggest that extreme ozone and extreme temperature are statistically significantly different in the original scale.

From our comparisons above, we have a summary for the comparisons among the three models as follows:

$$\text{Normalized scale} : SDM \approx_e GCM2000 \approx_e GCM2100,$$

and

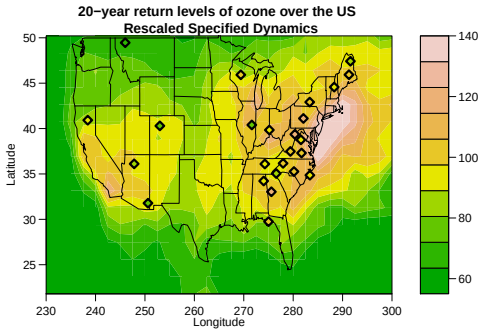
$$\text{Original scale} : SDM <_e GCM2000 <_e GCM2100.$$

The notations $<_e, \approx_e$ above are comparisons in the sense of extremes. These

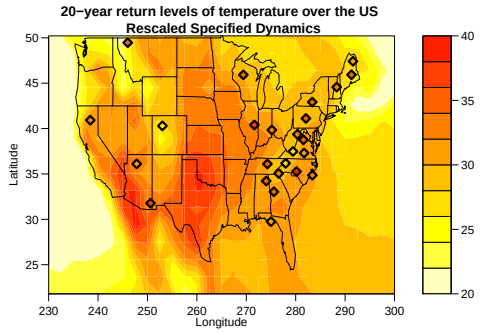
relationships hold for both extreme temperature and extreme ozone. However, this type of comparisons can only give information on the means of the attribute marginally. To evaluate the relationship jointly, we will use the tools we developed earlier in Section 4.3 to help us quantify the relationships between extreme temperature and extreme ozone. Before moving into that, we take a detour to visit marginal extremes to compare between CASTNET measurements and the models in the next section.

4.4.3 Comparison of Extreme Quantities

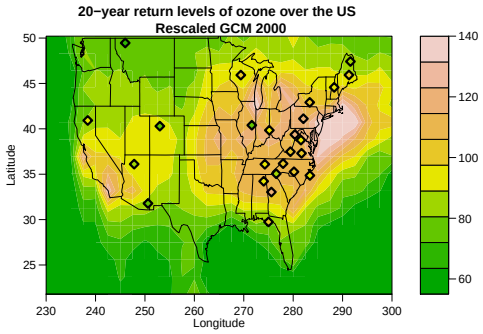
Recall the N-year return levels in (3.10) from the previous chapter. We provide 20-year return levels of both ozone and temperature from each model compared to those from CASTNET sites, see Figure 4.11. We use rescaled data in calculating these quantities (see (3.4) for the rescaling procedure). The 20-year return levels of ozone from CASTNET sites in eastern U.S. are slightly lower than those from the models, and the 20-year return levels of temperature from CASTNET sites are higher than those from SDM but they are about the same levels to GCMs. This suggests that the models effectively simulate marginal extremes for both ozone and temperature. Table 4.6 reports the 20-year return levels of temperature and ozone from each CASTNET site and the corresponding location from the models.



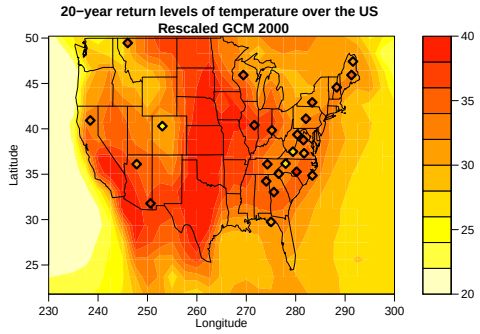
(a) Rescaled ozone from SDM



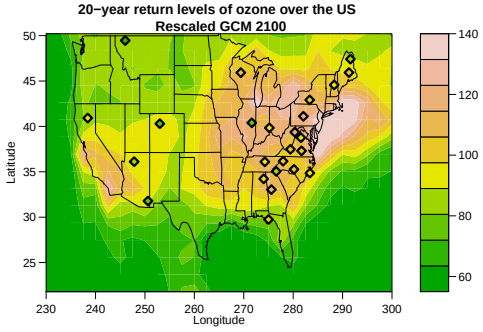
(b) Rescaled temperature from SDM



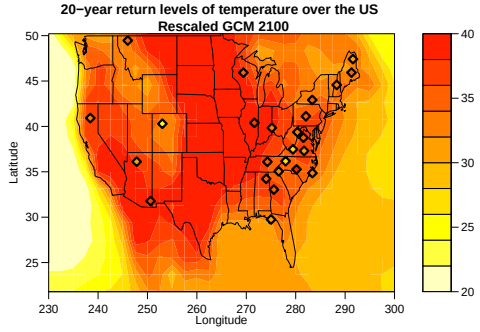
(c) Rescaled ozone from GCM 2000



(d) Rescaled temperature from GCM 2000



(e) Rescaled ozone from GCM 2100



(f) Rescaled temperature from GCM 2100

Figure 4.11: Left panels: 20-year return levels of ozone (ppb) for each model. Right panels: 20-year return levels of temperature ($^{\circ}C$) for each model. The diamonds are corresponding 20-year return levels from CASTNET sites. The plots use rescaled ozone and rescaled temperature in calculation to get rid of year-to-year effects and seasonal cycles. The rescale method is explained in (3.4).

Table 4.6: 20-year return levels of ozone (ppb) and temperature ($^{\circ}C$) from CASNTET sites and corresponding locations from each model (see Section 3.3.2 for the definition of return levels). The data is rescaled (see the rescaling procedure in (3.4)). For CASTNET data, we use the measurements from summertime 1992-2013. For each model, the data are summertime from different years: SDM = 1992-2010, GCM 2000 = 2006-2025, GCM 2100 = 2106-2125.

Site name	State	Ozone (ppb)				Temperature ($°C$)			
		CASTNET	SDM	GCM 2000	GCM 2100	CASTNET	SDM	GCM 2000	GCM 2100
Ashland	ME	87.48	98.72	86.31	92.65	31.85	26.85	32.29	32.45
Howland	ME	97.79	112.23	95.97	101.15	34.47	23.51	30.11	32.38
Woodstock	NH	93.31	98.36	95.83	100.75	32.75	28.12	32.12	32.9
Connecticut Hill	NY	103.28	103.77	106.87	112.71	34.57	28.33	30.43	33.68
Penn State	PA	115.78	104.16	116.37	118.97	35.02	29.32	32.56	35.99
Parsons	WV	114.7	103.7	104.14	105.25	33.55	28.13	32.46	35.7
Average	Northeast	102.06	103.49	100.92	105.25	33.7	27.38	31.66	33.85
Edgar Evins	TN	96.89	98.21	108.38	107.89	34.87	30.51	32.2	36.48
Georgia Station	GA	119.47	107.69	114.3	115.28	34.99	30.39	34.07	37.01
Sand Mountain	AL	91.86	94.31	112.53	114.95	34.78	29.45	33.17	35.15
Candor	NC	103.4	106.17	108.3	105.8	38.35	32.56	34.4	35.58
Coweeta	NC	85.16	92.71	107.31	109.04	33.37	29.05	31.98	34.91
Cranberry	NC	100.75	90.87	107.72	102.31	26.31	27.87	33.63	35.81
Beaufort	NC	96	106.36	108.47	100.6	33.94	31.96	34.65	35.34
Prince Edward	VA	98.7	109.95	109.14	111.63	31.46	31.54	34.83	37.46
Shenandoah NP	VA	92.03	121.79	119.21	116.48	35.68	32.21	33.78	36.42
Horton Station	VA	104.19	101.8	108.52	102.33	27.66	29.46	33.33	36.2
Sumatra	FL	100.96	76.28	75.7	75.44	30.45	30.68	31.26	31.75
Average	Southeast	99.04	100.56	107.23	105.61	32.9	30.52	33.39	35.65

Table 4.6: (Continued).

Site name	State	Ozone (ppb)				Temperature ($°C$)			
		CASTNET	SDM	GCM 2000	GCM 2100	CASTNET	SDM	GCM 2000	GCM 2100
Bondville	IL	75.31	104.87	114.93	111.27	36.17	31.27	37.24	40.98
Perkinstown	WI	104.14	97.68	96.57	110.57	34.97	30.66	36.02	40.56
Oxford	OH	90.55	120.16	119.08	120.66	32.23	32.1	35.32	38.5
Average	Midwest	90	107.57	110.19	114.17	34.46	31.34	36.19	40.01
Glacier NP	MT	65.33	71.36	75.92	82.01	35.05	23.03	27.49	32.94
Chiricahua	AZ	74.2	94.55	84.7	90.83	37.07	32.79	36.69	39.7
Grand Canyon	AZ	82.66	99.3	96.96	93.82	31.92	32.27	34.88	36.86
Gothic	CO	78.31	90.42	91.23	91.15	25.26	27.57	30.07	33.93
Lassen Volcanic	CA	92.54	84.7	80.37	80.19	32.78	31.16	33.67	36.83
Average	West	78.61	88.07	85.84	87.6	32.42	29.36	32.56	36.05
Average	All	94.59	99.6	101.79	102.95	33.18	29.63	33.15	35.82

4.4.4 Comparison in Ranks Method

In this section, we use ranks method described in Section 4.3.2 to help us quantifying extreme dependence between temperature and ozone. In order to use the ranks method, we use the temperature and ozone both in normalized scale unless stated.

First, we take a look at the scatter plots in Figure 4.12. The data is from a CASTNET site in Ashland (ME), and the data from models are from the corresponding grid points to this particular CASTNET site. We treat temperature and ozone as a vector (T, O_3). If we consider all points, we see that the scatter plots look similar but when we use ranks method to pick only extreme points (shown by red dots), they are different. The extreme measurements from CASTNET (shown by red points) seem to have a stronger relationship between ozone and temperature. Since we would like to quantify the extreme dependence between temperature and ozone, we apply ranks method to the data, estimate spectral measure, and calculate φ (see (4.18) for the definition of φ). The value of φ from this CASTNET site is higher than the values of φ from the models as well. The results are interesting enough, so we apply ranks method to all CASTNET sites and all grid points for each model.

Figure 4.13 depicts φ across the U.S. from different models: specified dynamics, GCM 2000, and GCM 2100. These models share some common traits of high φ values in southeastern U.S. However, in Midwest and western U.S., there are some variations among these models. The Specified Dynamics model tends to have higher φ in Midwest and western U.S. than the values of φ from GCMs. We also apply the ranks method to CASTNET data to measure the extremal dependence from real observations represented by diamonds to compare to each model; see

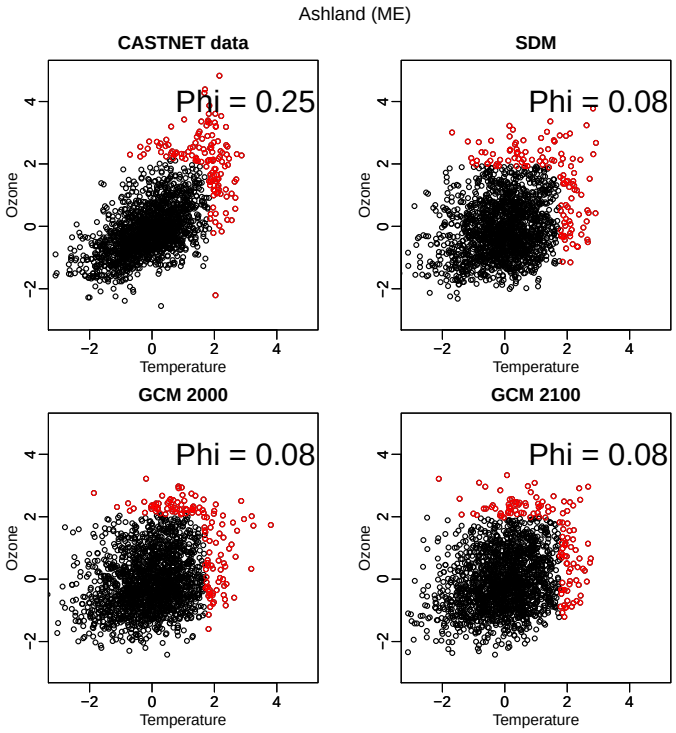
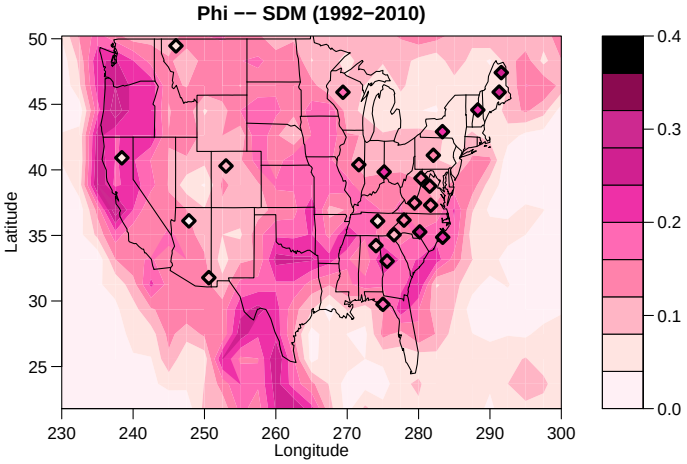
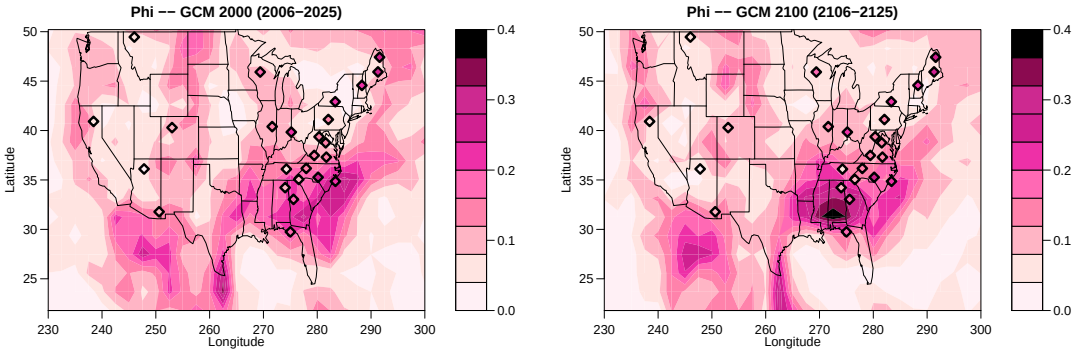


Figure 4.12: The scatter plots of temperature and ozone in the normalized scale from Ashland (ME) (or corresponding grid points to Ashland (ME)). The top left panel shows the normalized scale data from CASTNET; the top right panel shows the data from SDM; the bottom left panel shows the data from GCM 2000; and the bottom right panel shows the data from GCM 2100. Extreme points picked by the ranks methods are labeled in red (see the ranks method in Section 4.3.2).



(a) Comparisons between φ from CASTNET and SDM.



(b) Comparisons between φ from CASTNET and GCM 2000

(c) Comparisons between φ from CASTNET and GCM 2100

Figure 4.13: φ from CASTNET compared to (a) Specified Dynamics Model, (b) GCM 2000, and (c) GCM 2100. The values of φ from CASTNET are represented by diamonds and the values of φ from models are represented in the background. We use deseasonalized MDA8 ozone and deseasonalized daily maximum temperature before applying ranks method on the data.

Figure 4.13. We can see that most CASTNET sites in western U.S. have roughly the same values of φ as those from GCMs, while in northeastern U.S., φ's from CASTNET are higher than those from all the models. In contrast, φ from most of CASTNET sites in southeastern U.S. are about the same as φ from the models. Table 4.7 reports the values of φ from each model corresponding to each CASTNET site.

Table 4.7: φ of each model corresponding to CASTNET sites. For CASTNET data, we use the normalized measurements from 1992-2011 summertime. For each model, we use normalized summertime data with different years: SDM = 1992-2010, GCM 2000 = 2006-2025, GCM 2100 = 2106-2125.

Site name	State	CASTNET	SDM	GCM 2000	GCM 2100
Ashland	ME	0.25	0.08	0.08	0.08
Howland	ME	0.25	0.1	0.12	0.09
Woodstock	NH	0.21	0.04	0.05	0.05
Connecticut Hill	NY	0.22	0.09	0.07	0.06
Penn State	PA	0.15	0.08	0.04	0.05
Parsons	WV	0.14	0.13	0.07	0.12
Average	Northeast	0.2	0.09	0.07	0.08
Edgar Evins	TN	0.07	0.16	0.11	0.14
Georgia Station	GA	0.17	0.25	0.22	0.31
Sand Mountain	AL	0.09	0.16	0.23	0.2
Candor	NC	0.26	0.17	0.19	0.19
Coweeta	NC	0.11	0.14	0.18	0.13
Cranberry	NC	0.14	0.16	0.15	0.15
Beaufort	NC	0.2	0.19	0.26	0.15
Prince Edward	VA	0.16	0.13	0.17	0.19
Shenandoah NP	VA	0.15	0.09	0.04	0.11
Horton Station	VA	0.12	0.12	0.12	0.1
Sumatra	FL	0.18	0.12	0.2	0.14
Average	Southeast	0.15	0.15	0.17	0.16
Bondville	IL	0.14	0.1	0.1	0.08
Perkinstown	WI	0.17	0.1	0.1	0.07
Oxford	OH	0.24	0.15	0.1	0.07
Average	Midwest	0.18	0.12	0.1	0.07
Glacier NP	MT	0.08	0.13	0.03	0.04
Chiricahua	AZ	0.12	0.13	0.16	0.14
Grand Canyon	AZ	0.04	0.12	0.04	0.06
Gothic	CO	0.1	0.1	0.09	0.09
Lassen Volcanic	CA	0.08	0.2	0.14	0.08
Average	West	0.08	0.14	0.09	0.08
Average	All	0.15	0.13	0.12	0.12

We will take a closer look at selected CASNET sites: Ashland (ME), Sand Mountain (AL), Beaufort (NC). The reason we pick these sites is that they illustrate different characteristics for the differences of φ between CASTNET sites and the models: Ashland (ME) has φ from CASTNET higher than φ from the models; Sand Mountain (AL) has φ from CASTNET lower than φ from the models; and Beaufort (NC) has φ from CASTNET about the same level as φ from the models. Figures 4.12, Figure 4.14 and Figure 4.15 show scatter plots from CASTNET measurements, SDM, GCM 2000, and GCM 2100 from the selected CASTNET sites (or the corresponding grid point from particular model). The extreme points considered by the ranks methods are labeled in red. We may notice different levels of extreme dependence between temperature and ozone among CASTNET and models by considering only red points on each panel here. The plots from Ashland (ME) (Figure 4.12) show that extreme points from CASTNET site are away from the axes while some extreme points from the models are closer to axes than extreme points from CASTNET. The results are reverse if we consider the plots from Sand Mountain (AL). However, the plots from Beaufort (NC) are harder to tell the differences. Extreme points from CASTNET have slightly more points near the axes than SDM and GCM 2100. This is only a quick look at the red points to compare between CASTNET measurements and model data. φ helps us quantifying the differences. To compare the extreme dependence between models and CASTNET as a whole, we estimate φ for each sites of CASTNET data, call them $\{X_i; i = 1, ..., 25\}$. Then, for each model m ($m = 1$ refers to SDM, $m = 2$ refers to GCM 2000, and $m = 3$ refers to GCM 2100), we estimate φ from the grid points corresponding to CASTNET sites, call these $\{Y_{m,i}, i = 1, ..., 25\}$, and use the t-test to compare φ between the samples $\{X_i; i = 1, ..., 25\}$ and $\{Y_{m,i}, i = 1, ..., 25\}$. We compute the p-value under the null hypothesis for testing the equality of the means

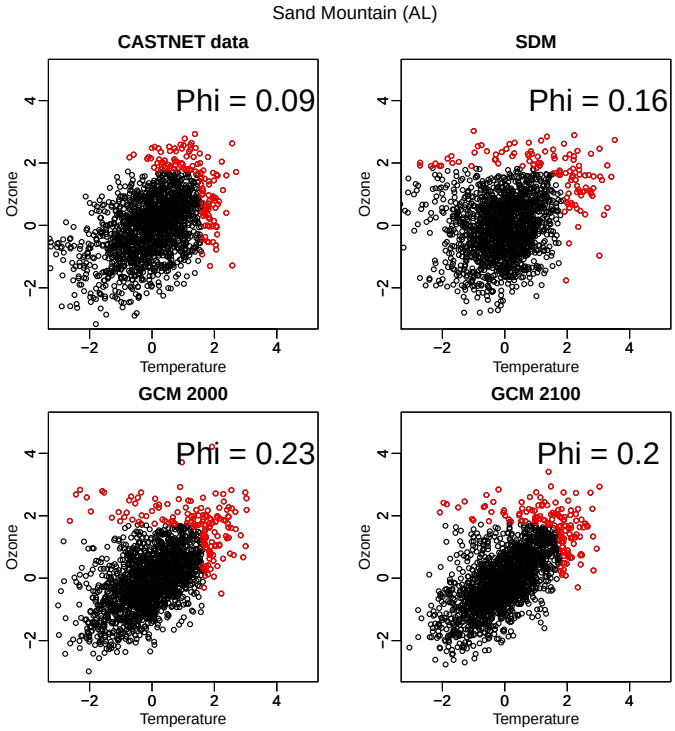


Figure 4.14: Similar to Figure 4.12 but we use the data from Sand Mountain (AL) (or corresponding grid points to Sand Mountain (AL)).

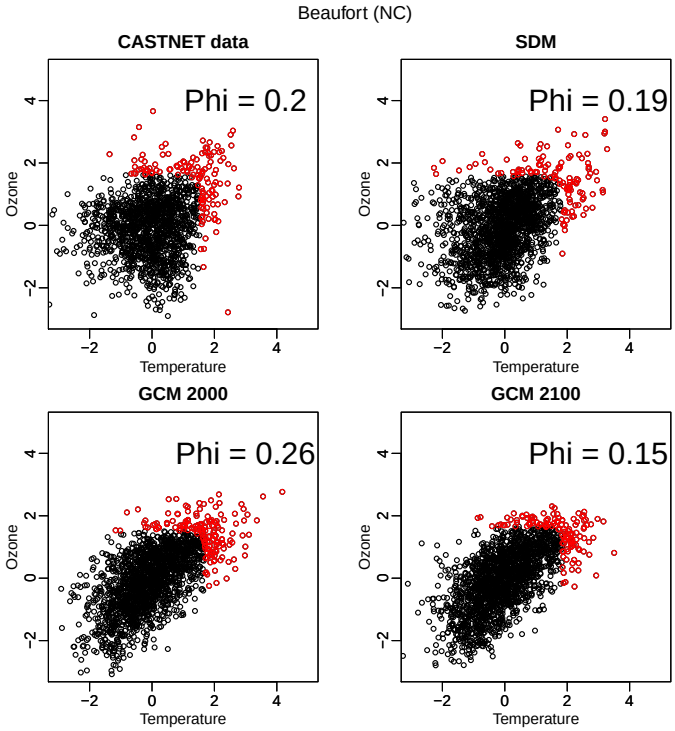


Figure 4.15: Similar to Figure 4.12 but we use the data from Beaufort (NC) (or corresponding grid points to Beaufort (NC)).

Table 4.8: P-values from the paired t-tests on the null hypothesis that a specific type of metric (column) from each model (row) is the same as from CASTNET data.

	p-value		
	φ	Shape Parameter	
Model		Ozone	Temperature
SDM	0.166	0.003	0.087
GCM 2000	0.042	0.051	0.290
GCM 2100	0.039	0.001	0.487

of $\{X_i\}$ and $\{Y_{m,i}\}$. In addition to φ, we repeat the process for ozone shape parameters and temperature shape parameters as well (see the description of shape parameter in Section 3.3.2). Table 4.8 reports the summary of the results from these tests. We conclude that we do not have enough evidence to reject the null hypothesis of that the temperature shape parameters from models are equal to the shape parameters from CASTNET data. However, for ozone shape parameters, we can reject the null hypothesis. That is, the models fail to represent ozone with the same shape parameter to CASTNET data. The study results as is because in each model, we use temperature as our independent variable while ozone is a dependent variable.

Lastly, the estimated angular measures have low p-values for the tests of CASTNET against GCMs and high enough p-value for the test against SDM (if we consider the 95% significance level). The estimated angular measure represents the extreme dependence between temperature and ozone and since GCMs fail to follow this dependency, we conclude that the models have flaws for simulating extremes. However, for SDM, we cannot conclude from our evidence if the model well represents the relationship between extreme temperature and extreme ozone.

4.5 Conclusion

In this study, we have explored the multivariate extreme value theory by adding temperature as another variable. To measure the extreme dependence between temperature and ozone, we prefer using spectral measures rather than correlations because correlations can only capture the linear relationship between the variables and according to many studies, as temperature increases, ozone stops acting linearly to temperature levels. However, a problem also arises if we want to use spectral measure because different variables are generally incomparable due to the differences in tail indices. We use the ranks method which essentially transforms data so that each component has the same tail index of 1, and then estimate the spectral measure to compare the tail dependence.

We applied the ranks method to the data generated by three different models– specified dynamics model (SDM), current general circulation model (GCM 2000), and future general circulation model (GCM 2100). The values of φ from SDM are slightly higher than those from GCMs. This indicates that SDM has higher extreme dependence than GCMs. Between the two GCM models, the values of φ are almost the same across the continental U.S. This shows the consistency in extreme dependence of the models that use the same simulation. However, if we compare the models to the observations from CASTNET sites, we have significant differences in terms of φ. In particular, CASTNET sites in Midwest and northeastern U.S. have higher φ than the model simulations. For western U.S., CASTNET sites have lower φ than the simulations. However, the results on the values of φ are dubious compared to the model simulations. Previous study conducted by Lamarque et al., (2012) [27] also reported that the specified dynamics model overestimated the observations from CASTNET in general. Although the

study did not have any conclusion in terms of extremes, it implies an excess of extreme dependence between temperature and ozone. The study from Shen et al., (2016)[44] also found out that when temperature is high enough, ozone stops rising according to temperature. This phenomena would refer to the low extreme dependence between temperature and ozone as well.

The discrepancies in spectral measure estimations (which imply discrepancies in extreme dependence) among the models and CASTNET sites suggest that the models do not well represent the real climate scheme, especially for extreme climate studies. The results suggest that if we have to use these models, we should use them with caution. As of now, we have to rely on the observations and possibly develop a new climate model that could represent extreme climate better.

CHAPTER 5

CONCLUSIONS

In this dissertation, we investigated several aspects of ozone and temperatures based on CASTNET measurements. (a) We evaluated the effects of NO_X SIP call on surface ozone levels. (b) We developed a procedure to measure the extreme behavior of ozone instead of using solely MLE. (c) We quantify extreme dependence between temperature and ozone and then compared this measure from CASTNET measurements to established climate models.

5.1 Changepoint Detection Framework

In Chapter 2, we use changepoint detection methods to find the most significant change in ozone and temperature means and variances of CASTNET measurements. For changepoint in means, we provided two methods: one is parametric method which assumes normality of the data, and the other one is non-parametric which is based on the rank of the data and this method can overcome the assumption of normality of the data. However, our study shows that both methods find quite the same changepoints from CASTNET sites. The procedure detects the significant changes of ozone means around 2002-2004 which agrees with the implementation of NO_X SIP call. For changes in temperature means, only few sites have the changes around the mid point of the time series, and the test statistics are not as significant as the changes in means. In fact, the temperature levels increase between the two periods.

5.2 Extreme Ozone Behavior

In Chapter 3, we investigated further into another aspect of the NO_X SIP call. Instead of considering the means of ozone distributions, we look into tail indices of ozone distributions. First, we propose to use Hill estimator to calculate the tail indices instead of using MLEs alone. We tested the procedure with synthetic data that are generated to be similar to the real ozone measurements to verify the procedure. It turns out that the effects of NO_X increases the shape parameters of ozone distributions. It could be even worse by changing the from negative shapes to positive ones which results in switching from bounded to unbounded upper limits. This raises more questions for the long term effects from the policy in the future as well.

5.3 Quantifying Extreme Dependence

In Chapter 4, we take temperature as an ozone covariate and add more data from three different models: specified dynamics model(SDM), general climate models for current climate (GCM 2000) and potential future climate (GCM 2100). The goal is to compare the CASTNET measurements and the models in extremes. We use the ranks method to normalize the tail indices so that we are able to quantify extreme dependence between temperature and ozone since the conventional methods, such as correlation, can only measure linear relationships. We start by comparing general quantities such as means, variances, and correlations. Next, we start imposing more conditions on the data with high temperature levels, and then, we compare the extreme dependence between CASTNET and the models. The results suggest that the models do not well represent extreme relationships

between temperature and ozone. A deeper analysis is needed for understanding the mechanisms of extreme ozone behavior.

5.4 Future Work

From Chapter 3, the limitations of the current procedure are that Hill estimator can only be used for positive shape parameter and we cannot guarantee when the results from either Hill estimator or MLE are between -0.3 and 0.3. A development of the procedure can by done by generalizing Hill estimator so that it can be used for estimating both positive or negative shape parameters.

From the results of Chapter 4, it shows that extreme ozone and extreme temperature need more understanding because from the current models, we tend to overestimate the relationship between them. We can extend the work from Chapter 4 to other variables from the models so that we understand extreme behavior of the climate better. Chapter 4 also suggests that the three models, which are actually widely used in climate studies, are far from realistically modeling extreme ozone and temperature. Future work here is to modify the climate models so that they can better reproduce extreme values as well.

BIBLIOGRAPHY

[1] N. Aleksic, J.-Y. Ku and L. Sedefian (2013): Effects of the NOx SIP Call program on ozone levels in New York. *Journal of the Air & Waste Management Association* 63:1335–1342.

[2] H. Alexandersson (1986): A homogeneity test applied to precipitation data. *Journal of Climatology* 6:661–675.

[3] D. E. Apollonio, N. Wolfe and L. A. Bero (2016): Realist review of policy intervention studies aimed at reducing exposures to environmental hazards in the United States. *BMC Public Health* 16:1–15.

[4] M. Bell, A. McDermott, S. Zeger, J. Samet and F. Dominici (2004): Ozone and short-term mortality in 95 US urban communities, 1987-2000. *Jama* 292(19):2372 – 2378.

[5] B. J. Bloomer, J. W. Stehr, C. A. Piety, R. J. Salawitch and R. R. Dickerson (2009): Observed relationships of ozone air pollution with temperature and emissions. *Geophysical Research Letters* 36:n/a–n/a. L09803.

[6] B. Brown-Steiner, P. Hess and M. Lin (2015): On the capabilities and limitations of GCCM simulations of summertime regional air quality: A diagnostic analysis of ozone and temperature simulations in the US using CESM CAM-Chem. *Atmospheric Environment* 101:134–148.

[7] T. J. Butler, F. M. Vermeylen, M. Rury, G. E. Likens, B. Lee, G. E. Bowker and L. McCluney (2011): Response of ozone and nitrate to stationary source {NOx} emission reductions in the eastern {USA}. *Atmospheric Environment* 45:1084 – 1094.

[8] S. Coles (2001): *An Introduction to Statistical Modeling of Extreme Values*. Springer, New York.

[9] L. de Haan and A. Ferreira (2006): *Extreme Value Theory: An Introduction*. Springer, New York.

[10] B. N. Duncan, Y. Yoshida, J. R. Olson, S. Sillman, R. V. Martin, L. Lamsal, Y. Hu, K. E. Pickering, C. Retscher, D. J. Allen and J. H. Crawford (2010): Application of {OMI} observations to a space-based indicator of {NOx} and {VOC} controls on surface ozone formation. *Atmospheric Environment* 44:2213 – 2223.

[11] D. R. Easterling and T. C. Peterson (1995): A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology* 15:369–377.

[12] J. H. Einmahl, V. I. Piterbarg and L. de Haan (2001): Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics* 29:1401–1423.

[13] P. Embrechts, C. Klüppelberg and T. Mikosch (1997): *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.

[14] C. Fang, R. Monson and E. Cowling (1996): Isoprene emission, photosynthesis, and growth in sweetgum (Liquidambar styraciflua) seedlings exposed to short- and long-term drying cycles. *Tree Physiology* 16:441–446.

[15] A. Fiore, J. Oberman, M. Lin, L. Zhang, O. Clifton, D. Jacob, V. Naik, L. Horowitz, J. Pinto and G. Milly (2014): Estimating North American background ozone in U.S. surface air with two independent global

models: Variability, uncertainties and recommendations. *Atmospheric Environment* 96:284–300.

[16] A. Guenther, X. Jiang, C. Heald, T. Sakulyanontvittaya, T. Duhl, L. Emmons and X. Wang (2012): The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *Geosci. Model Dev.* 5:1471–1492.

[17] A. Guenther, T. Karl, P. Harley, C. Wiedinmyer, P. Palmer and C. Geron (2006): Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmospheric Chemistry and Physics* 6:3181–3210.

[18] G. Guerova and N. Jones (2007): A global model study of ozone distributions during the August 2003 heat wave in Europe. *Environ. Chem.* 4:285–292.

[19] D. Hawkins and Q. Deng (2010): A nonparametric change-point control chart. *Journal of Quality Technology* 42:165–173.

[20] J. Hegarty, H. Mao and R. Talbot (2007): Synoptic controls on summertime surface ozone in the northeastern United States. *Journal of Geophysical Research: Atmospheres* 112:n/a–n/a. D14306.

[21] C. Hogrefe, B. Lynn, K. Civerolo, J.-Y. Ku, J. Rosenthal, C. Rosenzweig, R. Goldberg, S. Gaffin, K. Knowlton and P. Kinney (2004): Simulating changes in regional air pollution over the eastern United States due to changes in global and regional climate and emissions. *Journal of Geophysical Research: Atmospheres* 109:n/a–n/a. D22301.

[22] D. Horton, C. Skinner, D. Sngh and N. Diffenbaugh (2014): Occurrence and persistence of future atmospheric stagnation events. *Nature climate change* 4:698–703.

[23] IPCC (2013): *Summary for Policymakers*, book section SPM. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1–30.

[24] D. Jacob and D. Winner (2009): Effect of climate change on air quality. *Atmospheric Environment* 43:51–63.

[25] D. Jaffe and N. Wigder (2012): Ozone production from wildfires: A critical review. *Atmospheric Environment* 51:1–10.

[26] M. Jerrett, R. Burnett, C. Pope, K. Ito, G. Thurston, D. Krewski, Y. Shi, E. Calle and M. Thun (2009): Long-term ozone exposure and mortality. *The New England journal of medicine* 360(11):1085–1095.

[27] J.-F. Lamarque, L. K. Emmons, P. G. Hess, D. E. Kinnison, S. Tilmes, F. Vitt, C. L. Heald, E. A. Holland, P. H. Lauritzen, J. Neu, J. J. Orlando, P. J. Rasch and G. K. Tyndall (2012): CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model. *Geoscientific Model Development* 5:369–411.

[28] J.-T. Lin, K. Patten, K. Hayhoe, X.-Z. Liang and D. Wuebbles (2008): Effects of future climate and biogenic emissions changes on surface ozone over the United States and China. *Journal of Applied Meteorology and Climatology* 47:1888–1909.

[29] M. Lin, A. Fiore, O. Cooper, L. Horowitz, A. Langford, H. Levy, B. Johnson, V. Naik, S. Oltmans and C. Senff (2012): Springtime high surface ozone events over the western United States: Quantifying the role of stratospheric intrusions. *Journal of Geophysical Research: Atmospheres* 117:n/a–n/a. D00V22.

[30] J. Logan (1989): zone in rural areas of the United States. *ournal of Geophysical Research* 94:8511.

[31] L. Mickley, D. Jacob, B. Field and D. Rind (2004): Effects of future climate change on regional air pollution episodes in the United States. *Geophysical Research Letters* 31:n/a–n/a.

[32] A. M. Mood (1954): On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests. *Ann. Math. Statist.* 25:514–522.

[33] K. Murazaki and P. Hess (2006): How does climate change contribute to surface ozone change over the United States? *Journal of Geophysical Research: Atmospheres* 111:n/a–n/a. D05301.

[34] T. Nguyen and G. Samorodnitsky (2012): Tail Inference: where does the tail begin? *Extremes* 15:437–461.

[35] T. Nguyen and G. Samorodnitsky (2013): Multivariate tail estimation with application to analysis of CoVar. *ASTIN Bulletin* 43:245–270.

[36] W. R. Ott (1990): A Physical Explanation of the Lognormality of Pollutant Concentrations. *Journal of the Air & Waste Management Association* 40:1378–1383.

[37] J. Reeves, J. Chen, X. Wang, R. Lund and Q. Lu (2007): A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology* 46:900–915.

[38] S. Resnick (1987): *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.

[39] S. Resnick (2007): *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.

[40] H. Rieder, A. Fiore, L. Horowitz and V. Naik (2015): Projecting policy-relevant metrics for high summertime ozone pollution events over the eastern United States due to climate and emission changes during the 21st century. *Journal of Geophysical Research: Atmospheres* 120:784–800. 2014JD022303.

[41] H. Rieder, A. Fiore, L. Polvani, J.-F. Lamarque and Y. Fang (2013): Changes in the frequency and return level of high ozone pollution events over the eastern United States following emission controls. *Environ. Res. Lett.* 8: doi:10.1088/1748–9326/8/1/014012.

[42] G. J. Ross, D. K. Tasoulis and N. M. Adams (2011): Nonparametric Monitoring of Data Streams for Changes in Location and Scale. *Technometrics* 53:379–389.

[43] S. Russo, A. Dosio, R. Graversen, J. Sillmann, H. Carrao, M. Dunbar, A. Singleton, P. Montagna, P. Barbola and J. Vogt (2014): Magnitude of extreme heat waves in present climate and their projection in a warming world. *Journal of Geophysical Research: Atmospheres* 119:12500–12512.

[44] L. Shen, L. J. Mickley and E. Gilleland (2016): Impact of increasing heat waves on U.S. ozone episodes in the 2050s: Results from a multimodel analysis using extreme value theory. *Geophysical Research Letters* 43.

[45] D. Shindell, G. Faluvegi and N. Bell (2003): Preindustrial-to-present-day radiative forcing by tropospheric ozone from improved simulations with the GISS chemistry-climate GCM. *Atmospheric Chemistry and Physics* 3:1675–1702.

[46] S. Solberg, O. Hov, A. Sovde, I. Isaksen, P. Coddeville, H. de Backer, C. Forster, Y. Orsolini and K. Uhse (2008): European surface ozone in the extreme summer 2003. *Journal of Geophysical Research: Atmospheres* 113:n/a–n/a. D07307.

[47] A. L. Steiner, A. J. Davis, S. Sillman, R. C. Owen, A. M. Michalak and A. M. Fiore (2010): Observed suppression of ozone formation at extremely high temperatures due to chemical and biophysical feedbacks. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 107:19685–19690.

[48] D. Stevenson, F. Dentener, M. Schultz, K. Ellingsen, T. van Noije, O. Wild, G. Zeng, M. Amann, C. Atherton, N. Bell, D. Bergmann, I. Bey, T. Butler, J. Cofala, W. Collins, R. Derwent, R. Doherty, J. Drevet, H. Eskes, A. Fiore, M. Gauss, D. Hauglustaine, L. Horowitz, I. Isaksen, M. Krol, J.-F. Lamarque, M. Lawrence, V. Montanaro, J.-F. Muller, G. Pitari, M. Prather, J. Pyle, S. Rast, J. Rodriguez, M. Sanderson, N. Savage, D. Shindell, S. Strahan, K. Sudo and S. Szopa (2006): Mul-

timodel ensemble simulations of present-day and near-future tropospheric ozone. *Journal of Geophysical Research: Atmospheres* 111:n/a–n/a. D08301.

[49] E. Tagaris, K. Manomaiphiboon, K.-J. Liao, L. Leung, J.-H. Woo, S. He, P. Amar and A. Russell (2007): Impacts of global climate change and emissions on regional ozone and fine particulate matter concentrations over the United States. *Journal of Geophysical Research: Atmospheres* 112:n/a–n/a. D14312.

[50] R. van Dingenen, F. Dentener, F. Raes, M. Krol, L. Emberson and J. Cofala (2009): The global impact of ozone on agricultural crop yields under current and future air quality legislation. *Atmospheric Environment* 43(3):604–618.

[51] R. Vautard, M. Beekman, J. Desplat, A. Hodzic and S. Morel (2007): Air quality in Europe during the summer of 2003 as a prototype of air quality in a warmer climate. *Comptes Rendus Geoscience* 339:747–763.

[52] R. Vautard, C. Honore, M. Beekman and L. Rouil (2005): Simulation of ozone during the August 2003 heat wave and emission control scenarios. *Atmospheric Environment* 29:2957–2967.

[53] F. Vukovich (1995): Regional-scale boundary layer ozone variations in the eastern United States and their association with meteorological variations. *Atmospheric Environment* 29(17):2259–2273.

[54] A. Wilson, A. G. Rappold, L. M. Neas and B. J. Reich (2014): Modeling the effect of temperature on ozone-related mortality. *The Annals of Applied Statistics* 8:1728–1749.

[55] S. Wu, L. Mickley, D. Jacob, D. Rind and D. Streets (2008): Effects of 2000-2050 changes in climate and emissions on global tropospheric ozone and the policy-relevant background surface ozone in the United States. *Journal of Geophysical Research: Atmospheres* 113:n/a–n/a. D18312.